目　录

一、概 率 定 义

从本世纪三十年代起，概率论由一个很窄小的主题，发展成为一个宽广而与数学中许多其他领域有重要联系的分支．概率论是对客观存在的随机现象的数学抽象，正如几何学和力学一样，它同样也正确地反映了客观现实，具有广泛的应用．但是，概率论与几何学、力学在研究问题的方法上具有较大的差别，它所处理的不是决定性现象，而是随机现象，于是它有自己独特的概念和方法，在处理许多实际问题时具有独特的作用．

什么是“随机现象”呢？有这样一类现象，当人们对它加以观察或进行实验时，观察或实验的结果是许多可能结果中的某一个，而且这个或那个结果出现的机会是这个现象本身所固有的性质，我们把这种现象叫做**随机现象**，并把这些结果出现的机会的测度叫做**概率**．在初等水平上看这个问题，就是计算所出现的结果（叫做事件）的可能性大小．一个典型的例子是打桥牌，52 张牌分到四家，如果牌洗得彻底均匀，我们就可以计算某一家拿到 13 张不同点的牌的概率．计算这个概率，自然地会想到它应该等于下面这个比例：

$$\frac{\text{分 52 张牌时，某家拿到 13 张不同点的牌的可能情况的个数}}{\text{分 52 张牌时，这一家拿到 13 张牌的可能情况的个数}}.$$

接下来再求分子、分母的值，这只是计算个数的问题，属于组合数学的范畴了．由此可见，当我们谈概率时，总是把它结合着某种实

验来说的. 实验(或观察)的结果称为**事件**. 例如, 在掷两个骰子的实验中,"总和为 6 点"、"双 6"、"两个都是奇数点"都是可能掷出的结果, 所以它们都是事件. 在这个实验里, 最终可能的结果是: (1, 1), (1, 2), …, (1, 6), (2, 1), (2, 2), …, (2, 6), ……, (6, 1), (6, 2), …, (6, 6), 一共三十六种. 事件"双 6", 就是最后那个情形; 而事件"总和为六点", 是指出现(1, 5)或(2, 4)或(3, 3)或(4, 2)或(5, 1)中的任何一个. 由此可见, 事件可以区分为**简单事件**与**复合事件**. 例如"总和为 6 点"是个复合事件, 它是由 5 个简单事件构成的; 而"双 6" 本身就是一个简单事件.

对以上例子加以抽象, 我们看到: 简单事件即是该实验的所有各种可能结果(这些简单事件叫做**样本点**, 或简称为**点**); 复合事件可以分解为一些简单事件(也就是说, 复合事件是由多个点组成的集合). 实验中的每一个不再可分解的结果就是点; 由所有这些样本点构成一个总集合, 称之为**样本空间**. 由以上讨论可知, 当谈到概率论时, 首要的概念就是样本空间, 其中的点为简单事件, 这些点规定了该实验的所有可能结果, 而且只有从实验的每个可能结果(点)都能够清楚地判定某一事件 E 是发生或不发生时, 谈起 E 来才有确切的意义. 判定"E 发生了"这个事件的那些点, 构成一个集合, 它完全确定事件"E 发生了"; 反过来, 含一个或多个点的给定集合 E 都可以当作一个事件, 这个事件是否发生就由实验的结果(点)是否落在 E 中而定. 在概率中, 常说"事件 E 由某些点构成", 就意味着这些点便是在实验中能使 E 发生的那些结果.

1. 样 本 空 间

我们用 Ω 表示样本空间, 其中的点用 ω 表示; 事件是样本空间的子集合, 用 A、B、…… 等表示. 在一个实验中, 谈到事件时, 总是有不可能出现的事件, 这反映在子集合上就是不包含 Ω 中任何点的子集, 即**空集**, 用符号 $\varnothing$ 表示. 若 A 是一个事件, 则"A 不

发生"也是一个事件，它由 Ω 中 A 以外的诸点组成，记作 A^c，称为 A 的补事件．实际问题中提出的各种各样的事件及其相互关系，均可反映为样本空间 Ω 的点组成的各种子集合及其相互关系上，因而，可以把概率论的语言与集合论的语言相互对照如下：

事件及其相互关系	集合及其相互关系	符　　号
样本空间	空间	Ω
简单事件	点	ω
事件	集合	A, B, …
不可能事件	空集	$\varnothing$
必然事件	整个空间	Ω
A 发生蕴含 B 发生	A 是 B 的子集	$A\subset B$，包含运算
A 与 B 同时发生	A 与 B 之交	$A\cap B$ 或 AB，乘运算
A 与 B 至少其中之一发生	A 与 B 之并	$A\cup B$，并运算
A 发生但 B 不发生	A 与 B 之差	$A\setminus B$，差运算
A 不发生（A 的补事件）	A 的补集	A^c
A 与 B 不同时发生（不相容）	A 与 B 不交	$A\cap B=\varnothing$

事件间的关系，可以表为图 1.

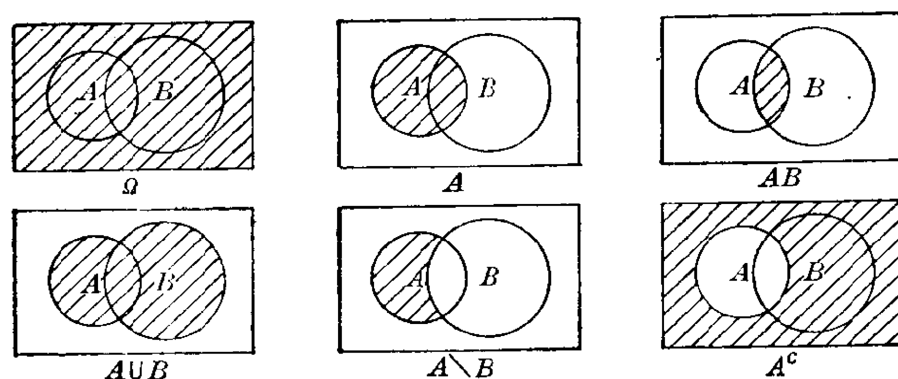


图 1

对于一串事件 A_1、A_2、A_3……，可定义两个新事件（即两个运算）：

(i) 属于全部给定集合的点所组成的集合，用 $\bigcap_{i=1}^{\infty} A_i$ 表示，即

$$\bigcap_{i=1}^{\infty} A_i = \{\omega \mid \omega \in A_i,\ i=1,2,\cdots\cdots\}.$$

或记 $\bigcap_{i=1}^{\infty} A_i = A_1A_2A_3\cdots\cdots$，称之为**交**．它表示 A_1、A_2、A_3、……诸事件同时发生；

(ii) 由至少属于给定集中之一的诸点所组成的集合，用 $\bigcup_{i=1}^{\infty} A_i$ 表示，即

$$\bigcup_{i=1}^{\infty} A_i = \{\omega \mid \omega \in A_i, \text{至少对某一个 } i\}.$$

或记 $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots$，称之为**并**．它表示事件 A_1、A_2、A_3、…中至少有一个发生．我们还可以用很明显的极限形式表达以上两个事件

$$\bigcup_{i=1}^{\infty} A_i = \lim_{n\to\infty} \bigcup_{i=1}^{n} A_i, \qquad \bigcap_{i=1}^{\infty} A_i = \lim_{n\to\infty} \bigcap_{i=1}^{n} A_i.$$

显见，当 $n\to\infty$ 时，$\bigcup_{i=1}^{n} A_i$ 不降 $\left(\text{即} \bigcup_{i=1}^{n} A_i \subseteq \bigcup_{i=1}^{n+1} A_i\right)$，而 $\bigcap_{i=1}^{n} A_i$ 不增．所以，前者涨到 $\bigcup_{i=1}^{\infty} A_i$，后者缩到 $\bigcap_{i=1}^{\infty} A_i$．

此外，对事件 (A_i) 为有穷个或无穷个事件时，都成立以下的运算：

$$(\bigcup_i A_i)B = \bigcup_i (A_iB),$$

$$(\bigcap_i A_i)^c = \bigcup_i A_i^c.$$

2. 古典概率

古典概率讨论的对象，限于实验所有可能结果为有穷个等可能的情形．这时，样本空间 Ω 由 n 个简单事件组成．若事件 E 由

m 个简单事件组成, 由直觉可以计算 E 的概率 $P(E)=\frac{m}{n}$, 这就是法国数学家拉普拉斯(P. S. M.de Laplace, 1749～1827)对古典概率作的定义. 让我们来计算打桥牌时, 分到西家一副 13 张不同点的牌之概率. 首先, 分到西家 13 张牌所有可能情况的副数等于 $C_{52}^{13}=\frac{52\cdot 51\cdot \cdots \cdot 40}{13!}=635013559600$ 副; 分到西家 13 张不同点的牌所有可能情况的个数即从 4 张 A 中抽一张, 4 张 K 中抽一张, ……, 所以总共可抽得 4^{13} 副不同点的牌. 从而, 所欲计算的概率等于

$$P=\frac{4^{13}}{C_{52}^{13}}\approx 0.0001057.$$

下面再举一个历史上有名的"得分问题". 甲、乙两人各出同样的赌注, 用扔硬币为赌博手段. 若正面朝上(记为"+"), 甲得 1 分, 乙不得分; 若反面朝上(记为"－"), 乙得 1 分, 甲不得分. 谁先得到事前约定的分数, 谁就赢得全部赌注. 当进行到甲还差两分, 乙还差三分才能达到约定分数时, 他们不愿继续赌下去, 问这时如何公平分配全部赌注? 解此问题, 显然, 为确保能分出胜负, 最多需要再扔 4 次, 这 4 次的所有可能结果为: {(++++), (+++－), (++－+), (++－－), (+－++), (－+++), (+－+－), (－++－), (－－－+), (－－－－), (+－－+), (－+－+), (－－++), (+－－－), (－+－－), (－－+－)}, 其中使甲获胜(即至少两个"+"的情形)有 11 种, 使乙获胜(即至少三个"－"的情形)有 5 种, 故甲胜的概率为 $\frac{11}{16}$, 乙胜的概率为 $\frac{5}{16}$, 因而甲应得全部赌注的 $\frac{11}{16}$, 乙应得 $\frac{5}{16}$. 实际上, 前四种情形表示只需扔两次就可分胜负, 往下的六种情形表示需扔三次即可分胜负, 最后的六种表示非要扔四次才能决定胜负. 但如果这样考虑, 就不利于用古典概率进行计算了.

计算古典概率, 可以像上例那样, 采用穷举法数清事件里面所

含简单事件的个数. 借助于组合计算, 可以简化计算过程. 如上例中, "甲胜"事件所含简单事件个数为 $C_4^4+C_4^3+C_4^2=11$. 需要强调的是,这里应十分注意"等可能"这个性质,即每个简单事件具有相同的概率. 历史上,法国数学家达朗贝尔(J. L. R. D'Alembert, 1717~1783)在考虑掷两个硬币的实验中, 曾错误地认为"出现一个正面朝上一个反面朝上"的概率为 $\frac{1}{3}$. 事实上, 此实验有 $\{(++), (+-), (-+), (--)\}$ 四个等可能的简单事件,而事件一反一正含 $(+-)$、$(-+)$ 两个简单事件,其概率应为 $\frac{1}{2}$.

在古典概率模型中,可把事件 A 的概率 $P(A)$ 看成事件 A 的函数,它具有下列三个性质:

(i) 对于任意事件 A, 有 $0\leqslant P(A)\leqslant 1$;

(ii) 必然事件 Ω 的概率 $P(\Omega)=1$;

(iii) 若 A、B 不相容, 其中至少一个发生即 $A\cup B$ (或记为 $A+B$, 对不相容的情形)的概率为 $P(A+B)=P(A)+P(B)$.

性质(iii)是计算概率的重要法则. 例如,有一批产品,总数为 1000 件, 其中 10 件为废品, 从这 1000 件中随意抽出 20 件, 求这 20 件中有废品的概率. 用 A 表示这一事件, 如直接计算 A 的概率, 就要依次计算 20 件中恰有 1 个、2 个、……直至 10 个为废品的概率, 再求这些概率之和. 但 A 的补事件 A^c 是这 20 件中无一废品. 由性质(ii)、(iii)知 $P(A)+P(A^c)=P(\Omega)=1$, 于是 $P(A)=1-P(A^c)$, 而计算 $P(A^c)$ 是很容易的, 它等于 $\frac{C_{990}^{20}}{C_{1000}^{20}}$, 故 $P(A)=1-\frac{C_{990}^{20}}{C_{1000}^{20}}$.

3. 几何概率

在建立古典概率的同时, 人们就注意到只考虑有穷个简单事件构成的样本空间是不够用的. 把等可能思想应用到含无穷个点

的样本空间，就产生了几何概率．其基本思想可以表述为：设 Ω 是平面上一个可求面积的区域，其面积记为 $\mu(\Omega)$．设 Ω 中的子集 A（即事件）的面积为 $\mu(A)$．随意扔一个点到 Ω 中，按照等可能性的思想，可以认为落入 Ω 中子集 A 的概率等于这个子集的面积与 $\mu(\Omega)$ 之比，即

$$P(A)=\frac{\mu(A)}{\mu(\Omega)}.$$

下面的蒲丰(G. L. de Buffon，1707～1788)投针问题是一个应用几何概率的典型例子．设平面上有一族平行线，每相邻两条之间的距离为 2 单位．取单位长的针一枚，随意把它扔到平面上，求针与直线相交的概率．用 x 表示针的中点到离它最近的一条平行线的距离，用 θ 表示针与此平行线的夹角（如图 2）．(x, θ) 完全决定针所落的位置．针的所有可能位置为 $0\leqslant x\leqslant 1, 0\leqslant\theta<\pi$（即图 3 所示的矩形里的每一个点），矩形的面积为 π．针与直线相交的充分必要条件是 $x\leqslant\frac{1}{2}\sin\theta$，这一不等式相应于图 3 中的阴影部分，其面积等于 $\frac{1}{2}\int_0^\pi\sin\theta\,d\theta=1$．用 A 表示事件“针与直线相交”，则 $P(A)=\frac{1}{\pi}$．利用这一公式，重复 n 次投针，计数它与直线相交的次数 m，则当 n 很大时，相交的频率 $\frac{m}{n}$ 可以作为 $P(A)=\frac{1}{\pi}$ 的近似（参看下节的讨论）．虽然这样做既费时，又不精确，但这一思想有可取之处，特别在有了电子计算机之后，可以用电子计算机模拟投针试验，计算 π 的近似值，这就显出其优越性了．许多与此

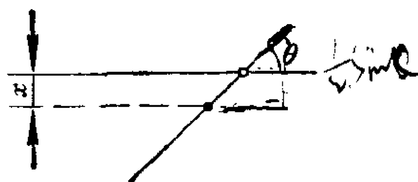


图 2

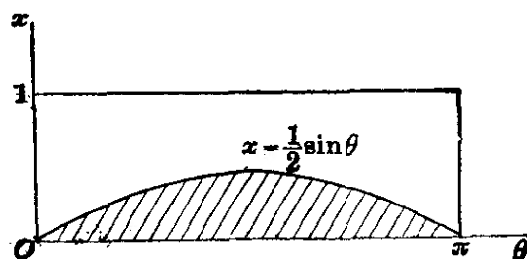


图 3

例相象的问题，表面上看来和概率毫不相干，但可用几何概率加以解决.

4. 概率的一般定义

数学科学的发展，总是结合着实际，把具体的对象扩展到抽象的对象，把特殊的事物推广到一般的事物. 首先我们来考虑等可能性的假设. 因为利用等可能性曾经遇到了悖论，从而引起人们注意到在一般情形下，等可能性这个前提是否必要. 下面先介绍这个悖论（贝特朗(Bertrand)悖论）:

在半径为 1 的圆内随意取一弦，计算此弦的长超过圆内接等边三角形边长 $\sqrt{3}$ 的概率.

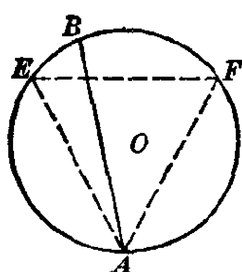


图 4

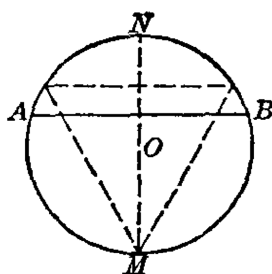


图 5

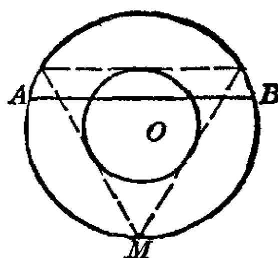


图 6

解此题有三种考虑方法:

如图 4，AB 为弦. 不失一般性，把弦的一端 A 固定，令点 B 在圆周上移动. 显然，当 B 点在 $\overset{\frown}{EF}$ 上时，AB 的长超过 $\sqrt{3}$，而 $\overset{\frown}{EF}$ 为圆周的 $\frac{1}{3}$，故所求概率为 $\frac{1}{3}$.

如图 5，弦 AB 的长与它与圆心 O 的距离有关，而与它的方向无关. 因此，可设它垂直于某个直径，不妨设此直径为 MN. 由图 5 看出，当且仅当弦与圆心的距离小于 $\frac{1}{2}$ 时，弦的长大于 $\sqrt{3}$. 因此，所求概率为 $\frac{1}{2}$.

如图 6，弦的长被它的中点所决定，从图中可以看出，当且仅

当中点在半径为 $\frac{1}{2}$ 的同心圆内时，弦长大于 $\sqrt{3}$. 但这个同心圆的面积为大圆面积的 $\frac{1}{4}$, 故所求概率为 $\frac{1}{4}$.

这个悖论的根源，在于取弦时采用了不同的等可能性，图 4 是把 B 点在圆周上看成有等可能性，图 5 是把弦中点在直径上看成有等可能性，而图 6 则把弦的中点在圆内看成有等可能性. 由此例看出，等可能性引起了怪现象!

在一些具体的例子如掷骰子实验中，人们还对事物加以理想化，如假设骰子为一个绝对对称的正六面体. 但是，现实中这种纯而又纯的事物是不存在的. 更深一步，即使存在绝对对称的骰子，但由于各人的掷法不同，也会破坏等可能性. 此外，我们对每个点出现的概率规定为 $\frac{1}{6}$, 这是出于人们的常识. 但对常识需要有直观的解释，$\frac{1}{6}$ 应该给出每个点大概要出现的机会的一个度量，而这个度量在经验上与多次掷骰子时某个点出现的频率相联系. 更一般地，令 E 表示"所出现的结果小于 5 点"，设我们重复掷 n 次骰子，令 $N_n(E)$ 表示这 n 次结果中 E 出现的次数，于是 E 出现的频率 $Q_n(E)=N_n(E)/n$, 我们有理由取 $Q_n(E)$ 作为事件 E 出现的度量. 当然 $Q_n(E)$ 与 n 有关，而且当 n 增多时，$Q_n(E)$ 不断地在摆动，甚至摆动很大. 但若 n 趋于无穷，随着 n 的增大，序列 $Q_n(E)$ 是否能在一个固定值的很小范围内摆动呢? 也就是 $Q_n(E)$ 是否稳定到一个固定值呢? 深究起来，我们不能答覆此问题，因为这是一个永不完结的实验. 但在数学上，我们可以理想化而规定这个极限存在，并记其极限为 $Q(E)$，这就是 $Q(E)=\lim\limits_{n\to\infty}Q_n(E)$，并称 $Q(E)$ 为事件 E 的直觉极限频率. 但这个极限，更深究起来，仍然与所取的具体的实验序列有关，即对第一个人作的实验(假定可以作)与第二个人，第三个人……所作的实验来说，不能保证他们能得到相同极限. 然而，从实际的直觉，我们需要有一个衡量事件 E 出现的机会的度量，而这个度量将比只作一次 $n\to\infty$ 的实验所得

的记录要含有更多的东西. 一个能够维持下去的以频率为基础的理论需要假定:对一切"相类似"的序列而言,上述极限都是一样的等于$Q(E)$. 但这一假设里,"相类似"的含义还是不明确的. 所以,这种办法只在一定范围内可以有效,而且,在直觉的常识范围内它具有诱惑力. 这里,我们不去穷追此问题. 下面将对事件E的概率给以合理的定义,在此定义下所得到的概率理论中可以证明一个基本定理:在某些条件下,事件E的频率$\frac{N_n(E)}{n}$,对实际所有能考虑到的序列而言,它的极限是存在的,且等于事件E的概率. 这个定理就是**大数定理**,它是一切实验科学应用概率论的基础,而且,在理论上验证了以频率来解释概率的直觉观念.

设Ω为样本空间,其中的事件具有样本空间里关于事件间的一切关系及运算. 我们定义每个事件E的概率$P(E)$为0与1之间的实数,它度量E发生的机会. 这种度量,犹如力学中的质量的度量或几何学中距离的度量,是事件所固有的性质. 至于这些度量等于什么,怎样测出来的,用不着再作其它假定. 但当人们真需要知道概率数值时,就需要人们的经验及统计方法来定出. 为了演算和推理,概率或称概率测度应具有初等概率所具有的三个性质,我们把这三个性质作为规定(公理):

(i) 对每个事件$E(\subset\Omega)$, $P(E)\geqslant 0$;

(ii) 对任意两个不相容事件E_1、E_2,它们的并(即至少有一个发生的事件)E_1+E_2的概率满足:

$$P(E_1+E_2)=P(E_1)+P(E_2);$$

(iii) $P(\Omega)=1$.

由以上三条公理,可以推出以下的性质:

(iv) 对任意事件E, $P(E)\leqslant 1$.

这是因为,E与其补事件E^c之并等于样本空间Ω,即$E+E^c=\Omega$;于是,由(ii)知$P(E)+P(E^c)=P(\Omega)=1$,从而,由(i)知$P(E)=1-P(E^c)\leqslant 1$.

(v) 对任意两个事件 E_1、E_2，且 $E_1\subset E_2$，则有

$$P(E_1)\leqslant P(E_2),\quad P(E_2\setminus E_1)=P(E_2)-P(E_1).$$

这是因为$E_2=E_1+(E_2\setminus E_1)$，于是 $P(E_2)=P(E_1)+P(E_2\setminus E_1)\geqslant P(E_1)$.

(vi) 对任意有穷个事件 E_1、E_2、…、E_n 且两两不相容，有下列的有穷可加性：

$$P(E_1+E_2+\cdots+E_n)=P(E_1)+P(E_2)+\cdots+P(E_n).$$

(vii) 对任意有穷个事件 E_1、E_2、…、E_n，恒成立下列布尔不等式：

$$P(E_1\cup E_2\cup\cdots\cup E_n)\leqslant P(E_1)+\cdots+P(E_n).$$

这是因为，当 $n=2$ 时，$E_1\cup E_2=E_1\cup E_1^cE_2$，对右边可用公理(ii)，得 $P(E_1\cup E_2)=P(E_1)+P(E_1^cE_2)$；但 $E_1^cE_2\subset E_2$，再用公理(v)，即得 $P(E_1\cup E_2)\leqslant P(E_1)+P(E_2)$. 对一般的 n，可用数学归纳法来证.

(viii) 对任意两个事件 E_1、E_2，有

$$P(E_1\cup E_2)+P(E_1\cap E_2)=P(E_1)+P(E_2).$$

这是因为，$P(E_1\cup E_2)=P(E_1)+P(E_1^cE_2)$，但 $E_1^cE_2=E_2\setminus E_1\cap E_2$，故有 $P(E_1\cup E_2)=P(E_1)+P(E_2)-P(E_1\cap E_2)$.

为了应付更广泛的情形，我们将把公理(ii)加以强化. 这种强化是符合客观实际的，而且由它推得的结论也是合理的. 公理(ii)及其直接推论(vi)可强化为**“可列可加性”**公理：

(ii*) 对两两不相容的事件的无穷序列 E_i(其中 $i=1,2,\cdots$)，有

$$P\left(\bigcup_{i=1}^{\infty}E_i\right)=\sum_{i=1}^{\infty}P(E_i).$$

注意，$\bigcup\limits_{i=1}^{\infty}E_i$ 仍是事件，它是诸 E_i (其中 $i=1, 2, \cdots$)中至少一个发生的事件.

如果 Ω 为有穷个点构成的，则(ii*)化为(ii). 但是，重要的是可列可加性不能由有穷可加性(vi)令 $n\to\infty$ 而得到. 现在让我们

看一看：有穷可加性可写为

$$P\left(\bigcup_{i=1}^{n} E_i\right)=\sum_{i=1}^{n} P(E_i), \quad E_i \text{ 两两不相容}(i=1, \cdots, n).$$

当 n 增加时，右方是非降的，而且其和不超过 1，故右方有极限 $\lim\limits_{n\to\infty}\sum\limits_{i=1}^{n} P(E_i)=\sum\limits_{i=1}^{\infty} P(E_i)$. 于是，由上式得

$$\lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} E_i\right)=\sum_{i=1}^{\infty} P(E_i).$$

欲使此等式化为可列可加性，需成立下列关系：

$$\lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} E_i\right)=P\left(\bigcup_{i=1}^{\infty} E_i\right)=P\left(\lim_{n\to\infty} \bigcup_{i=1}^{n} E_i\right).$$

但是，此式两端需要运算"$\lim\limits_{n\to\infty}$"与"P"可以互换. 如果读者学过微积分，就会知道这种互换不是随便可以进行的，而是有条件的.

二、基 本 性 质

1. 事件的独立性

独立性是概率论特有的概念. 考虑"甲扔硬币，乙掷骰子"的实验，求事件"硬币出现正面、并且骰子出现 5 或 6 点"的概率，若把"硬币出现正面"记作事件 A，"骰子出现 5 或 6 点"记作事件 B，则欲求的概率就是事件 $A\cap B$ 的概率. 整个实验的简单事件的总个数为 12，而且是等可能的. A 所含简单事件的个数为 6，B 所含简单事件的个数为 4（即，(正, 5)，(反, 5)，(正, 6)，(反, 6)），而 $A\cap B$ 所含简单事件的个数为 2，于是 $P(A)=\frac{6}{12}=\frac{1}{2}$，$P(B)=\frac{4}{12}=\frac{1}{3}$，$P(A\cap B)=\frac{2}{12}=\frac{1}{6}$. 由此可知 $P(A\cap B)=P(A)P(B)$. 对于任意两个事件 A 和 B，如果它们各自发生的概

率与它们同时发生的概率满足 $P(A\cap B)=P(A)P(B)$，则称它们为相互独立的. 结合上例，独立性这一定义与直观是一致的，因为甲和乙作实验是单独进行而互不影响的.

对于 n 个事件 A_1、A_2、…、A_n，$n\geqslant 2$，如果对所有可能的 $1\leqslant i_1<i_2<\cdots<i_l\leqslant n$（其中 $2\leqslant l\leqslant n$），都成立

$$P\left(\bigcap_{j=1}^{l} A_{i_j}\right)=\prod_{j=1}^{l} P(A_{i_j}),$$

则称这 n 个事件相互独立. 以上的关系式需要对 $l=2$，…，n 和任意 $1\leqslant i_1<i_2<\cdots<i_l\leqslant n$ 都成立，缺一不可. 这可由下面的例子说明：A、B、C 三个事件是两两相互独立，但非相互独立. 考虑一个正四面体，一面为红色，一面为蓝色，一面为黄色，剩下的那面为红、蓝、黄三色. 以 A、B、C 分别表示掷四面体时，朝下的一面出现红、蓝、黄的事件. 于是 $P(A)=P(B)=P(C)=\frac{1}{2}$，且 $P(AB)=P(AC)=P(BC)=\frac{1}{4}$. 由此知 A、B、C 两两相互独立. 但 $P(ABC)=\frac{1}{4}$，$P(A)P(B)P(C)=\frac{1}{8}$，不满足 $n=l=3$ 所要求的条件，所以事件 A、B、C 不相互独立.

[例] （**伯努利**(Jacob Bernoulli, 1654～1705)**实验**）对相同条件下重复的同一实验，独立性是很有用的. 例如重复扔硬币 n 次，$n\geqslant 2$，则整个重复试验的结果是由"正面"、"反面"组成的长为 n 的序列，我们常常把"正"、"反"分别量化为"1"，"0"，或"1"、"-1". 现采取前面一种量化，于是重复扔硬币 n 次的实验结果是由"0"、"1"组成的长为 n 的序列. 例如 $n=9$，(110010100) 就是一个结果. 对重复扔 n 次，一共有 2^n 个这样的序列. 由对称性假设及独立性，每一个简单事件即出现某一"0"、"1" n 序列的概率为 $\frac{1}{2^n}$. 如果把对称性假设去掉（有这类的实际问题），并设出现 1 的概率为 $p\left(\neq\frac{1}{2}\right)$，而出现 0 的概率自然是 $1-p$（记为 q），现在用数学公式来表示，设 X_i 为这 n 次重复试验中的第 i 次的结果，

于是

$$P(X_i=1)=p, \quad P(X_i=0)=1-p=q, \quad i=1, \cdots, n.$$

令 ε_i 为等于 0 或等于 1 的数 $(i=1, 2, \cdots, n)$，于是对任一串特定的 ε_1、ε_2、$\cdots$、ε_n，由独立性有

$$P(X_1=\varepsilon_1, X_2=\varepsilon_2, \cdots, X_n=\varepsilon_n)$$
$$=P(X_1=\varepsilon_1)P(X_2=\varepsilon_2)\cdots P(X_n=\varepsilon_n).$$

此式右方每项或等于 p，或等于 q. 如果诸 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 中有 j 个为 1，$n-j$ 个为 0，则上等式的右方等于 $p^j q^{n-j}$. 注意到对于重复扔 n 次的实验，$\sum_{i=1}^{n} X_i$ 表示出现正面的次数，但 $p^j q^{n-j}$ 不等于"恰出现 j 个正面"的概率，而只是"出现一个特定序列，其中恰有 j 个正面"的概率，例如"前 j 个为正面以后的都是反面"的概率. 为了计算"恰出现 j 个正面"的概率，我们必须数一数共有几个特定的序列，其中恰有 j 个正面. 易知这个数目等于 $C_n^j=\frac{n!}{j!(n-j)!}$. 如果令 $S_n=\sum_{i=1}^{n} X_i$，那么 $\{S_n=j\}$ 就是事件"恰出现 j 个正面". 故由概率的可加性，可知

$$P\{S_n=j\}=\sum P\{X_1=\varepsilon_1, X_2=\varepsilon_2, \cdots, X_n=\varepsilon_n\}$$
$$=\sum P(X_1=\varepsilon_1)P(X_2=\varepsilon_2)\cdots P(X_n=\varepsilon_n)$$
$$=C_n^j p^j(1-p)^{n-j}, \qquad j=0, 1, \cdots, n$$

这个公式叫作伯努利公式. S_n 的分布 $C_n^j p^j(1-p)^{n-j}=C_n^j p^j q^{n-j}$ 叫做二项式分布.

2. 随机变量及分布函数

以扔硬币一次为例，我们把样本点"正"与"反"分别量化为 1 与 0. 扔一次的结果用 $X(\omega)$ 表示，当 $\omega=$"正"时，$X(\omega)=1$；$\omega=$"反"时，$X(\omega)=0$. 这时 $X(\omega)$ 就是随机变量. 又如上面的例子中的 S_n，也是随机变量. 其中的 ω 为 $(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)$，而

$S_n(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)=j$ 当且仅当$(\varepsilon_1, \cdots, \varepsilon_n)$中在特定的 j 个位置上为1，其余的为0. 又如 $\Omega=\{\omega_1, \cdots, \omega_n\}$ 为 n 个人所组成的样本空间(以 ω_i 对人编号)，于是随意抽查其中一人的身高和体重，就可用两个随机变量 X 和 Y 表示，它们分别是 Ω 上的函数: $X(\omega_k)=$"ω_k 的身高"，$Y(\omega_k)=$"ω_k 的体重"，$k=1, 2, \cdots, n$. 一般说来，一个随机变量 X 是一个定义于样本空间上的函数，但这个函数有随机性，即它依赖于样本点的随机性，X 所取的值可以是离散的(如掷骰子的点数只取1到6的整数，电话台收到的呼叫次数只取非负整数)，也可以在一个数值空间或整个实数轴上取值.

在研究随机变量的性质时，确定和计算它取某个数值或落入某个数值区间内的概率，是特别重要的. 因此，使随机变量取某个数值或落入某个数值区间的简单事件的集合，应当构成实验的事件. 根据这样的要求，取实数值的随机变量的数学定义是: X 为 Ω 上的实值函数，且对任意实数 x，使 $X(\omega)\leqslant x$ 成立的一切 ω，即 $\{\omega|X(\omega)\leqslant x\}$ 是事件，它常简记为 $\{X\leqslant x\}$.

在一些场合要同时考虑多个随机变量. 例如气体分子运动时，可用 $X(\omega)$ 与 $Y(\omega)$ 分别表示一个气体分子的水平与垂直速度(因为大量分子相互碰撞，一个气体分子的运动是随机性的)，于是 $\sqrt{X^2+Y^2}$ 表示这个分子的绝对速度，这也是随机变量. 粗略地说，若 $f(x, y)$ 为 x、y 的二元函数，具有一定的光滑性，X、Y 为随机变量，则 $f(X, Y)$ 也是随机变量，例如 $X\pm Y$、XY、$\frac{X}{Y}$ (其中 $Y\neq 0$) 以及它们的线性组合 $aX+bY$ 都是随机变量.

现在我们再看看日常生活里是如何用到随机变量的. 常常是预先直觉地知道某些变量具有随机性. 事实上，人们在讨论随机变量 X、Y 等等时，往往并不必为确定样本空间 Ω 所困扰. 形式的数学结构只供必要逻辑推演的幕后背景，但不必每次用到概率语言时都把数学构架拉到前台. 例如以图书印刷为例，印数在

1000 本以内每本成本为 3 元，从 1000 本以上到 5000 本时，这数目中的每本书的成本减至 2 元，超过 5000 本的每本书的成本减为 1 元．厂家所能假定的是如何定出售价，书店在最初印 1000 册，而且定为每本 5 元．在这行为中，随机性表现在出售的册数．这时如果能出售 1000 册，他就赚 2000 元；但有积压时，也不致于大赔本．令 X 为出售册数，Y 表示厂商的赔赚，于是，X 是随机的，所以 Y 也是随机变量，而且它们之间的关系如下：

$$Y=\begin{cases}5X-3000, & \text{当 } X\leqslant 1000 \text{ 时；}\\ 2000+3(X-1000), & \text{当 } 1000<X\leqslant 5000 \text{ 时；}\\ 14000+4(X-5000), & \text{当 } X>5000 \text{ 时．}\end{cases}$$

根据以上关系，我们可提出出版图书赔钱的概率是多少，这就是计算下列事件的概率：

$$\{5X-3000<0\}=\{X<600\}.$$

同样可以提出图书出版者至少赚 10000 元的概率是多少，这就是计算下列事件的概率：

$$\begin{aligned}&\{2000+3(X-1000)\geqslant 10000\}\cup\{X>5000\}\\ &=\left\{X\geqslant\frac{8000}{3}+1000\right\}\cup\{X>5000\}\\ &=\{X\geqslant 3667\}\cup\{X>5000\}\\ &=\{X\geqslant 3667\}.\end{aligned}$$

因为 $\{X>5000\}\subset\{X\geqslant 3667\}$，计算上述事件的概率，就分别化为求事件 $\{X<600\}$ 及 $\{X\geqslant 3667\}$ 的概率，亦即事件 $\{X\leqslant 599\}$ 及 $\{X>3666\}$ 的概率（因为 X 取非负整数）．由此可见，对一个随机变量 X 来说，对每个实数 x，知道事件 $\{X\leqslant x\}$ 的概率 $P\{X\leqslant x\}$ 就是了解了 X 的规律，我们称 $P\{X\leqslant x\}=F(x)$ 为随机变量 X 的**分布函数**．由于 X 的取值范围不同，X 的分布函数分为**离散型**的与**连续型**的．例如电话台每天收到呼叫的次数 X 是取值于 $(0, 1, 2, \cdots)$ 中的随机变量，这时 X 的概率可以写为

$$P\{X=k\}=p_k, \qquad \sum_{k=0}^{\infty} p_k=1;$$

而分布函数为

$$F(x)=\sum_{k\leqslant x} p_k=P\{X\leqslant x\}.$$

它的图形为如图 7 所示的阶梯函数，在每个 k 处有一个跳跃，跳跃的高度为 p_k.

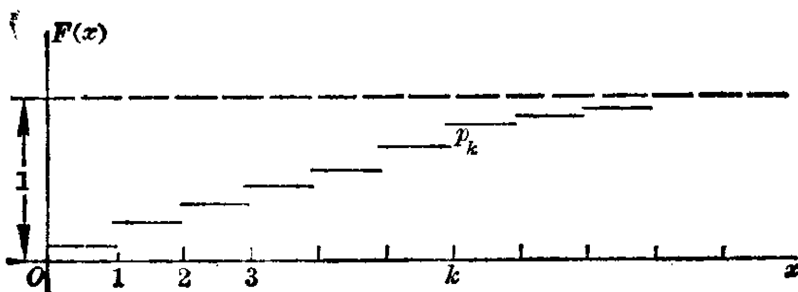


图 7

如果随机变量 X 取值为整个实轴，那么它的分布函数

$$F(x)=P\{X\leqslant x\}, \qquad -\infty<x<\infty$$

是非降的函数，而且

$$\lim_{x\to-\infty} F(x)=\lim_{x\to-\infty} P\{X\leqslant x\}=0,$$

$$\lim_{x\to\infty} F(x)=+1.$$

这时 $F(x)$ 的图形一般形状为如下图形：

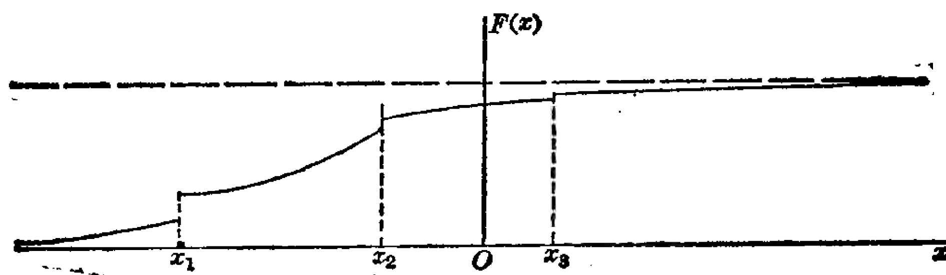


图 8

这里，分布函数可以在可列个 (x_i) 上有跳跃，其跳跃的高度分别为

$$F(x_i)-F(x_i-),$$

此处 $F(x_i-)=\lim_{x\to x_i} F(x)$，而且对任意 $a<b$，$P(a<X\leqslant b)=$

$F(b)-F(a)$.

常见的分布函数是具有概率密度的分布函数，即其概率密度

$$p(x)=\frac{d}{dx}F(x),$$

这时必有

$$F(x)=P(X\leqslant x)=\int_{-\infty}^{x}p(u)\,du \quad \left(\int_{-\infty}^{\infty}p(u)\,du=1\right),$$

$$P(a<X\leqslant b)=\int_{a}^{b}p(u)\,du,$$

$$P(x<X\leqslant x+\Delta x)=\int_{x}^{x+\Delta x}p(u)\,du\approx p(x)\Delta x.$$

密度函数 $p(x)$ 为非负函数，而且它的图形与 x 轴之间所围的面积为 1.

常见的概率密度为正态分布：

$$y=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad \sigma>0,$$

其图形为钟形，以直线 $y=m$ 为对称轴(图 9). 常把此分布记为 $\mathcal{N}(m,\ \sigma^2)$.

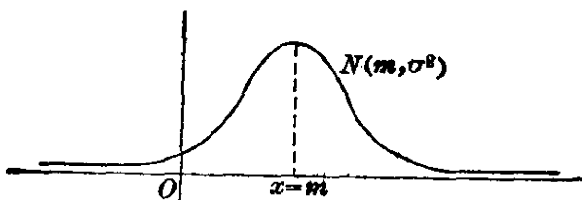


图 9

如果同时考虑多个随机变量，就称为多元随机变量，从而形成多维分布函数及密度函数的概念. 为简便计，只以二元情形为例. 这时有两个随机变量 X、Y. 如果 X 和 Y 的取值范围都是离散的，分别为 x_i(其中 $i=0,\ 1,\ 2,\ \cdots$) 及 y_j(其中 $j=0,\ 1,\ 2,\ \cdots$)，于是$(X,\ Y)$ 的二元分布为

$$P\{X=x_i,\ Y=y_j\}=p_{ij}, \qquad \text{其中 } i、j=0,1,2,\cdots.$$

由此可以计算与$(X,\ Y)$有关的事件的概率，特别是求单个 X 的分布时，则需使 Y 取遍所有 y_i (其中 $i=0,\ 1,\ 2,\ \cdots$)，这就是 X 的边际分布

$$P(X=x_i)=\sum_{j=0}^{\infty}p_{ij}.$$

对 Y 的边际分布，也同样有：$P(Y=y_k)=\sum_{i=0}^{\infty}p_{ik}$. 要注意只知各分量的边际分布一般说来不能求得$(X, Y)$的二元分布(联合分布).

对取值于整个二维空间的二元随机变量(X, Y)，我们定义二元联合分布为

$$P(X\leqslant x,\ Y\leqslant y)=F(x, y), \qquad -\infty<x、y<\infty.$$

这个分布具有下列性质，令 $\lim\limits_{y\to\infty}F(x, y)=F(x, \infty)$，则有

$$\begin{aligned}F(x, \infty)&=P\{X\leqslant x,\ Y<\infty\}\\&=P\{X\leqslant x\},\end{aligned}$$

因为"$Y<\infty$"实际上是对 Y 不加限制. 作为 X 的函数，$F(x, \infty)$是 X 的边缘分布函数. 同样，Y 的边缘分布函数为 $F(\infty, y)=\lim\limits_{x\to\infty}F(x, y)$.

常用的二元分布（连续值的）是具有概率密度 $p(x, y)$ 的分布函数：

$$0\leqslant p(x, y)=\frac{\partial^2}{\partial x\,\partial y}F(x, y),$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(u, v)dudv=1,$$

$$P(X\leqslant x,\ Y\leqslant y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u, v)dudv,$$

$$P(Y\leqslant y)=\int_{-\infty}^{y}\int_{-\infty}^{\infty}p(u, v)dudv,$$

$$\begin{aligned}P(X\leqslant x)&=\int_{-\infty}^{\infty}\int_{-\infty}^{x}p(u, v)dudv\\&=\int_{-\infty}^{x}du\int_{-\infty}^{\infty}p(u, v)dv,\end{aligned}$$

$$P(X\in[x,\ x+\Delta x],\ Y\in[y,\ y+\Delta y])\approx p(x, y)\Delta x\Delta y.$$

对于随机变量 X、Y，如果由它们所引起的事件是相互独立

的，那么我们称这二个随机变量为**独立随机变量**. 对取离散值的随机变量，“独立性”意味着对一切可能值 x_i、y_j，有 $P\{X=x_i, Y=y_j\}=P(X=x_i)P(Y=y_j)$；对取连续值的情形，独立性表现为：对一切实数 x、y，有 $P\{X\leqslant x, Y\leqslant y\}=P(X\leqslant x)P(Y\leqslant y)$. 对任意有穷个随机变量 X_1、X_2、…、X_n，也可以定义它们的相互独立性，这和事件的独立性完全相类似.

3. 数学期望

对随机变量而言，只要完全知道它的分布函数，我们就知道了与此随机变量相应的随机现象的规律，从而可利用此规律解决实际问题. 但是，有些与分布有关的量也是很有用的. 随机变量 X 的**数学期望**(或称**期望**，或称**均值**)是分布的很重要特征量，它的定义来自习惯上的“平均”这个概念. 假定由家庭所构成的样本空间(总体)中，有 k 个小孩的家庭有 n_k 个，其中 $k=0$、1、2、……，于是家庭的总数为 $n=n_0+n_1+n_2+\cdots\cdots$，而小孩的总数为 $m=n_1+2n_2+3n_3+\cdots$. 于是，每个家庭的小孩平均数为 $\frac{m}{n}=\frac{n_1}{n}+2\frac{n_2}{n}+3\frac{n_3}{n}+\cdots$. 由概率与频率的相似性的启发，我们给出“期望”的如下定义：X 的数学期望 EX 定义为

$$EX=\sum_{\omega\in\Omega}X(\omega)P\{(\omega)\}=\sum_{i=1}^{\infty}X(\omega_i)P\{(\omega_i)\}.$$

这里 Ω 为离散的样本空间，即 $\Omega=(\omega_1, \omega_2, \cdots)$，当然需要此级数是收敛的，即需要**绝对收敛**：

$$\sum_{\omega\in\Omega}|X(\omega)|P(\{\omega\})<\infty,$$

这时，我们称 X 的“期望”存在. 为了计算 EX，我们给出如下的方法. 设 Ω 可以分解为不相交的子集 A_n：即 $\Omega=\bigcup_n A_n$ 使得 $X(\omega)$ 在 A_n 取同一值 a_n，即 $X(\omega)=a_n$ 当 $\omega\in A_n$ 时，这里 a_n 可以不必全不相等. 就有

$$EX=\sum_n a_n P(A_n).$$

这个式子是把定义式中的 ω 首先划归成子集 A_n，然后再对 n 求和. 特别要指出的是，定义式还包含 X 的一些函数的期望:

$$E\varphi(X)=\sum_\omega \varphi(X(\omega))P(\{\omega\})=\sum_n \varphi(a_n)P(A_n).$$

特别，取 $\varphi(x)=|x|^r$，$E|X|^r$ 叫做 X 的 r 阶绝对矩. $X-EX$ 仍是一个随机变量，它的二阶矩 $E(X-EX)^2=EX^2-(EX)^2$ 叫做 X 的方差，记作 $\sigma^2(X)$.

设 X_1、…、X_n 都是具有有穷期望值的随机变量，则它们的和的期望存在，且有 $E(X_1+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)$. 只需对两个随机变量加以证明. 应用前面的符号知

$$\begin{aligned}EX+EY&=\sum_{j,k} x_j P(X=x_j,\quad Y=y_k)\\&\quad+\sum_{j,k} y_k P(X=x_j,\ Y=y_k),\end{aligned}$$

求和是对 (x_j, y_k) 跑遍一切可能的值. 由于右边的第一项和第二项都绝对收敛，从而得出 $\sum\limits_{j,k}(x_j+y_k)P(X=x_j,\ Y=y_k)$，这就等于 $E(X+Y)$. 不难证明，如果 X、Y 为相互独立的随机变量，则 $E(XY)=EXEY$.

[例 1] 重复扔硬币直到正面出现. 令 X 表示直到正面出现所扔的次数，$\{X=n\}$ 就意味着扔到第 n 次时才出现正面，而前 $n-1$ 次都出现的是反面. 这时由独立性知

$$\begin{aligned}p_n\triangleq P(X=n)&=P(\underbrace{\text{反}\cdots\cdots\text{反正}}_{n})\\&=\underbrace{P(\text{反})\cdots P(\text{反})P(\text{正})}_{n}=\frac{1}{2^n},\end{aligned}$$

于是

$$EX=\sum_{n=1}^{\infty}\frac{n}{2^n}=2.$$

这和直观相符，平均说扔二次即可出现正面. 如果硬币不是理想的，即 $P(\text{正})=p\left(\neq\frac{1}{2}\right)$，$P(\text{反})=1-p=q$，这时 $p_n=P\{X=$

$n\} = q^{n-1}p$, 从而有

$$EX = \sum_{n=1}^{\infty} npq^{n-1} = p\sum_{n=0}^{\infty}(n+1)q^n = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

随机变量 X 叫做出现正面的**等待时间**，如果把正面解释为"成功"，则 X 叫做"成功"的等待时间，X 的分布为 p_n（其中 $n=1, 2, \cdots$），$p_n = q^{n-1}p$，此分布叫做以"成功"概率为 p 的**几何分布**.

仍考虑伯努利实验. 令 S_n 为重复扔 n 次非对称硬币时出现正面的次数，这时第 i 次扔硬币的结果是随机变量 X_i，于是 $S_n = X_1 + \cdots + X_n$，而且

$$P(S_n = k) = C_n^k p^k q^{n-k}, \qquad k = 0, 1, \cdots, n.$$

由此知 $ES_n = \sum_{k=0}^{n} kC_n^k p^k q^{n-k} = np$. 还可利用求期望的加法公式计算：因为 $EX_i = p$（其中 $i=1, \cdots, n$），故有

$$ES_n = E(X_1 + \cdots + X_n) = EX_1 + \cdots + EX_n = np.$$

[例 2] 盒里盛有 N 个标有号码 1 到 N 的球. 随意取出一个，记其号码，放回去；再随意取一个记其号码，连续作 n 次，令 X 表示这 n 次抽取所记录的 n 个号码中的最大数. 事件 $\{X \leqslant k\}$ 表示这 n 个记录中每个号码都不大于 k（其中 $k = 1, 2, \cdots, N$），故 $P\{X \leqslant k\} = \left(\frac{k}{N}\right)^n$，而且

$$P\{X = k\} = P\{X \leqslant k\} - P\{X \leqslant k-1\} = \frac{k^n - (k-1)^n}{N^n},$$

由此可知

$$EX = \sum_{k=1}^{N} kP(X = k) = N^{-n}\left\{N^{k+1} - \sum_{k=1}^{N}(k-1)^n\right\}.$$

当 n 相当大时，$\sum_{k=1}^{N}(k-1)^n$ 约等于 $N^{n+1}/n+1$，故有

$$EX \approx \frac{n}{n+1}N.$$

这个式子很有用. 例如一个城市有一万辆汽车，其牌号由 1 到 10000. 我们随意观察 100 辆，由上可知记录的最大号码的期望

$EX\approx\frac{100}{101}10000\approx9901$，这个数可以作为 10000 的估计．二次世界大战时，曾用此法估计敌方生产某种产品的总数．

上面所讲的，是离散样本空间中随机变数的数学期望．如果样本空间不是离散的，随机变量 X 的分布为 $F(x)$，具有概率密度 $p(x)$，随机变量 X 的期望应如何计算呢？可用以下的粗略方法计算．任取很密的一组数：

$$-\infty<x_1<x_2<\cdots<x_n<\infty \quad (n \text{ 很大})$$

于是

$$P(x_i<X\leqslant x_{i+1})=p(x_i)(x_{i+1}-x_i) \qquad \text{（有密度情形）}$$

$$=F(x_{i+1})-F(x_i). \qquad \text{（一般分布的情形）}$$

X 的期望可用下式逼近

$$\sum_{i=1}^{n}x_ip(x_i)(x_{i+1}-x_i) \qquad \text{（有密度情形）}$$

或

$$\sum_{i=1}^{n}x_i(F(x_{i+1})-F(x_i)). \qquad \text{（一般分布的情形）}$$

当 n 增大，而且相邻的分点 x_i、x_{i+1} 之间的距离越来越小，上列两个和分别趋于：

$$\int_{-\infty}^{\infty}xp(x)dx, \qquad \int_{-\infty}^{\infty}xdF(x),$$

如果这两个积分存在．从下面的比较表，容易看出各种情形之间的相似状况：

	离 散 情 形	有密度情形	一般情形
样本空间	$\omega_i,\ i=1,2,\cdots$	$-\infty<x<\infty$	$-\infty<x<\infty$
$P(a<X\leqslant b)$	$\sum_{a<a_n\leqslant b}P(X(\omega)=a_n)$	$\int_a^b p(u)du$	$F(b)-F(a)$
$P(X\leqslant x)$	$\sum_{a_n\leqslant x}P(X(\omega)=a_n)$	$\int_{-\infty}^{x}p(u)du$	$F(x)$
EX	$\sum_n a_nP(X(\omega)=a_n)$	$\int_{-\infty}^{\infty}up(u)du$	$\int_{-\infty}^{\infty}udF(u)$
限制条件	$\sum_n\|a_n\|P(X(\omega)=a_n)<\infty$	$\int\|u\|p(u)du<\infty$	$\int_{-\infty}^{\infty}\|u\|kF(u)<\infty$

此外，对一些函数 φ, $\varphi(X)$ 的期望定义为

$$E\varphi(X)=\int_{-\infty}^{\infty}\varphi(u)p(u)du \quad \text{或} \quad \int_{-\infty}^{\infty}\varphi(u)dF(u),$$

如果积分存在的话.

[例 3] 电话台的每两次相邻的呼叫之间的时间间隔是一个随机变量. 很长的公路上一个观察站观察相继两辆车过此站的时间间隔也是一个随机变量. 这类随机变量叫做等待时间，记作 T. 在解决打电话难或车辆拥挤的现象时，等待时间 T 及其期望 ET 是很重要的参数，根据经验，一般取等待时间 T 的分布为**指数分布**，其参数为 $\alpha>0$, 即其分布密度 $p(x)$ 为

$$p(x)=\begin{cases}\alpha e^{-\alpha x}, & \text{当 } x\geqslant 0 \text{ 时;}\\ 0, & \text{当 } x<0 \text{ 时.}\end{cases}$$

由此知分布函数为

$$P\{T\leqslant t\}=F(t)=\int_0^t p(x)dx=1-e^{-\alpha t},$$

$$ET=\int_0^{\infty}\alpha u e^{-\alpha u}du=\frac{1}{\alpha}.$$

由此可知，α 愈小，平均时间间隔愈长. 平均时间间隔愈短表示交通愈繁忙，参数 α 叫作强度. 强度大，表示繁忙程度大.

4. 条件概率及条件期望

想计算独立事件同时发生或独立随机变量分布时，求各自的概率或各自的分布的乘积即可. 但对不独立的事件，计算同时发生的概率就需要**条件概率**，设一个城市有 N 个男人和 M 个女人，其中患色盲者男性为 n 人，女性为 m 人. 用 Ω 表示全体居民的集合，其中的点为 $M+N$ 个，A 表示女性居民的集合，其中的点为 M 个，B 表示全体色盲患者的集合，其中的点为 $m+n$ 个. 某厂要招一名化验员，要求不是色盲者，问任意一位应招的女性，被厂方拒绝的概率多大？这里要求的就是事件"女性中色盲患者"的概率，

用记号表示为 $P(B|A)$. 按等可能性原则,

$$P(B|A)=\frac{m}{M}=\frac{m/M+N}{M/M+N}=\frac{P(AB)}{P(A)}$$

或 $P(AB)=P(B|A)P(A)$. 把此例推广到任意两个事件 A 及 B 上, 我们定义: **已知事件 A 发生时, B 发生的条件概率为**

$$P(B|A)=\frac{P(AB)}{P(A)},\quad 这里\ P(A)>0.$$

这个条件概率简称为"已知 A 时, B 的条件概率"或"B 关于 A 的条件概率". 当 $P(A)=0$ 时有 $P(AB)=0$, 这时 $P(B|A)$成为 $\frac{0}{0}$, 此时没有意义(实际上也无用), 所以当我们用到条件概率时, 总想到 $P(A)>0$. 例如, 在重复扔硬币的实验中, 已知前三次出现反面, 问再扔二次而出现正面的概率是多少? 如本书第 21 页中的例 1, 含 X 为出现正面的等待时间, 则所要求的条件概率为

$$P(X\leqslant 5|X\geqslant 4)=\frac{P(4\leqslant X\leqslant 5)}{P(X\geqslant 4)}.$$

已经算过 $P(X=n)=pq^{n-1}$(其中 $n=1, 2, \cdots$), 由此知

$$P(X\geqslant 4)=\sum_{n=4}^{\infty}pq^{n-1}=\frac{q^3p}{1-q}=q^3,$$

$$P(4\leqslant X\leqslant 5)=q^3p+q^4p,$$

故
$$P(X\leqslant 5|X\geqslant 4)=p+qp.$$

此外还有 $P(1\leqslant X\leqslant 2)=p+qp$. 比较这两个等式的右方, 可知前三次出现反面不影响今后出现正面的等待时间的概率, 这现象可以事前想象得到, 但这一结果是由于每次重复扔硬币中的各次实验是相互独立的.

已知条件 A 时, 任意事件的条件概率可记作 $P(\cdot|A)$, "$\cdot$"可以取任何事件, 这样一来 $P(\cdot|A)$满足概率的公理(i)、(ii)、(iii), 即 $P(B|A)\geqslant 0$, 对任意事件 B; 若 $E_1\cap E_2=\varnothing$, 则 $P(E_1+E_2|A)=P(E_1|A)+P(E_2|A)$; $P(\Omega|A)=1$. 如果对某个 B, $P(B|A)=P(B)$, 于是有 $P(AB)=P(A)P(B|A)=P(A)P(B)$, 这表明

A、B 相互独立．关系式 $P(B|A)=P(B)$ 可解释为不论有没有"已知 A"的条件，B 的概率不变，这也就是 A、B 相互独立的直观意义。

关于条件概率，有三个重要公式：

(i) 一般乘法公式　设 E_1、E_2、…、E_n 为事件，$n\geqslant 2$，$P(E_1E_2\cdots E_{n-1})>0$，则有

$$P(E_1E_2\cdots E_n)=P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1E_2\cdots E_{n-1}).$$

把这个公式应用到随机变量 X_1、X_2、…、X_n 上，可得

$$\begin{aligned}&P\{X_1\leqslant x_1,\ \cdots,\ X_n\leqslant x_n\}\\&=P\{X_1\leqslant x_1\}P\{X_2\leqslant x_2|X_1\leqslant x_1\}P\{X_3\leqslant x_3|X_1\leqslant x_1,\\&\quad X_2\leqslant x_2\}\cdots P\{X_n\leqslant x_n|X_1\leqslant x_1,\ \cdots,\ X_{n-1}\leqslant x_{n-}\ \}.\end{aligned}$$

(ii) 全概率公式　设 E_1、E_2、… 为事件序列，两两不相容，$P\left(\bigcup\limits_{i=1}^{\infty}E_i\right)=1$，且对一切 i 有 $P(E_i)>0$，则对任意事件 E，有

$$P(E)=\sum_{i=1}^{\infty}P(E|E_i)P(E_i).$$

(iii) 贝叶斯公式　在(ii)的条件下，对任意正概率事件 E，有

$$P(E_n|E)=\frac{P(E|E_n)P(E_n)}{\sum\limits_{i=1}^{\infty}P(E_i)P(E|E_i)}.$$

贝叶斯公式可以计算"反概率"，即基于观察到的后果事件 E 计算出"原因" E_n 的概率，这就是通常所谓的"执果循因"．此时 $P(E_n)$ 叫做先验概率，而 $P(E_n|E)$ 为原因 E_n 的后验概率．贝叶斯公式很有用，例如 E 是一件凶杀事件，而 E_n 为许多被怀疑的凶手($n=1,\ 2,\ \cdots$)，于是由先验 $P(E_n)$ 及 $P(E|E_n)$ 可以计算 $P(E_n|E)$，从而判定 E_n 中哪个后验概率最大，它就是怀疑最大的人．

条件期望定义为对条件分布所取的数学期望．设 X、Y 为随机变量，它们的联合分布为 $F(x,\ y)=P\{X\leqslant x,\ Y\leqslant y\}$，欲求当

已知 $x<X\leqslant x+\Delta x$ 时，Y 的条件分布的期望时，先求条件分布

$$P\{Y\leqslant y\,|\,x<X\leqslant x+\Delta x\}=\frac{P\{x<X\leqslant x+\Delta x,\ Y\leqslant y\}}{P\{x<X\leqslant x+\Delta x\}}$$

$$=\frac{\int_{-\infty}^{y}\int_{x}^{x+\Delta x}dF(u,\ v)}{\int_{-\infty}^{\infty}\int_{x}^{x+\Delta x}dF(u,\ v)},$$

于是条件期望为

$$E(Y\,|\,x<X\leqslant x+\Delta x)=\frac{\int_{-\infty}^{\infty}\int_{x}^{x+\Delta x}v\,dF(u,\ v)}{\int_{-\infty}^{\infty}\int_{x}^{x+\Delta x}dF(u,\ v)}.$$

下面再举历史上有名的**配对问题**．这是应用条件概率的一个典型例子．设有 n 个信封，n 张信纸，分别编号为 1 到 n．把信纸随意装进信封里，一个信封恰装一张信纸，求“没有一个对上号”的概率 P_0，以及“恰有 r 个对上号”的概率 P_r，这里 $r=1,\ 2,\ \cdots,\ n$.

用 A_i 表示事件“第 i 张信纸恰装入第 i 个信封”，则 $P(A_i)=\frac{1}{n}$．在 A_i 发生的条件下，第 $j(\neq i)$ 张信纸只有 $n-1$ 个信封供它选择，故 $P(A_j|A_i)=\frac{1}{n-1}$，因此 $P(A_iA_j)=P(A_j|A_i)P(A_i)=\frac{1}{n(n-1)}$．类似地，对 $1\leqslant i_1<i_2<\cdots<i_r\leqslant n$，有 $P(A_{i_1}\cdots A_{i_r})=\frac{1}{n(n-1)\cdots(m-r+1)}$，从而

$$\sum_{1\leqslant i_1<\cdots<i_r\leqslant n}P(A_{i_1}\cdots A_{i_r})=C_n^r\frac{1}{n(n-1)\cdots(n-r+1)}=\frac{1}{r!}.$$

根据逐步淘汰原则，至少有一个对上号的概率

$$P\Big(\bigcup_{i=1}^{n}A_i\Big)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\cdots$$

$$+(-1)^{r-1}\sum_{1\leqslant i_1<\cdots i_r\leqslant n}P(A_{i_1}\cdots A_{i_r})+\cdots$$

$$+(-1)^{n-1}P(A_1A_2\cdots A_n)$$

$$=\sum_{r=1}^{n}\frac{(-1)^{r-1}}{r!},$$

从而 $P_0=1-P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{r=0}^{n}\frac{(-1)^r}{r!}$.

再计算 P_r. 由于对指定的某 r 张信纸，它们都对上号的概率为 $\frac{1}{n(n-1)\cdots(n-r+1)}$，余下的 $n-r$ 张信纸没有一张对上号的概率为 $\sum_{i=0}^{n-1}\frac{(-1)^i}{i!}$ (即在 P_0 中以 $n-r$ 代替 n). 而 r 张信纸对上号共有 C_n^r 种选法, 故

$$P_r=C_n^r\frac{1}{n(n-1)\cdots(n-r+1)}\sum_{i=0}^{n-r}\frac{(-1)^i}{i!}$$
$$=\frac{1}{r!}\sum_{i=0}^{n-r}\frac{(-1)^i}{i!}.$$

如果令 $n\to\infty$, 则 $P_r\to\frac{1}{r!}e^{-1}$, $r=0, 1, 2, \cdots$. 这个分布是参数为 1 的普阿松分布(参看本书第 31 页的极限定理).

三、极 限 理 论

1. 中心极限定理

先以重复扔对称硬币为例. 重复扔十次, 出现 0 次, 1 次, ……乃至 10 次正面的概率容易计算如下:

$$p_k=\frac{C_{10}^k}{2^{10}}, \qquad k=0, 1, \cdots, 10.$$

我们得到

p_0	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
0.001	0.01	0.045	0.12	0.21	0.25	0.21	0.12	0.045	0.01	0.001

把它们画成直方图(图 10). 这个图以 5.5 为对称轴. 如果我们把重复扔的次数增加到一万次再画直方图, 横坐标有 10001 段, 而对称轴在 5000.5 处, 最高直方约等于 $\frac{1}{100\sqrt{\pi}}=0.0056$. 但当我们把直方图所有高度放大 $\frac{100}{\sqrt{2}}$ 倍, 同时把横底缩小 $\frac{100}{\sqrt{2}}$ 倍, 于是所有直方的顶端几乎可以联成一条连续曲线, 这个曲线的方程是下列曲线的平移:

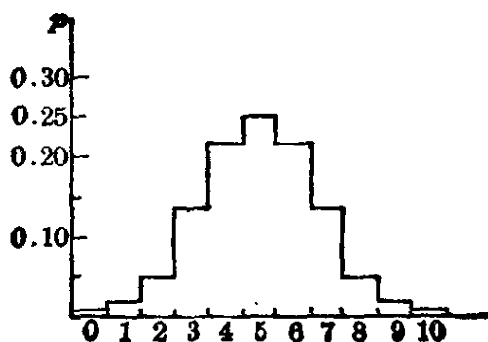


图 10

$$y=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

此曲线是 $m=0$、$\sigma^2=1$ 的正态曲线, 叫做标准正态曲线.

用数学语言来表述以上的现象. 不妨考虑非对称的硬币, 第 i 次扔硬币出现的结果用随机变量 X_i 表示, $X_i=1$ 表示出现正面, $X_i=0$ 表示出现反面, 且 $P(X_i=1)=p\left(\neq\frac{1}{2}\right)$, $P(X_i=0)=1-p=q$. 令 $S_n=X_1+\cdots+X_n=\sum_{i=1}^{n}X_i$, 它表示扔 n 次时正面出现的次数. 易知期望 $EX_i=p$, 方差 $\sigma^2(X_i)=E(X_i-EX_i)^2=pq$. 由独立性, 可算出 $ES_n=np$, $\sigma^2(S_n)=E(S_n-ES_n)^2=npq$. 用以上所讲的概率计算法则和斯特林(J. Stirling)公式, 可以推出著名的**德莫阿-拉普拉斯中心极限定理**: 对任意常数 a、b, $-\infty<a<b<\infty$, 有

$$\lim_{n\to\infty}P\left\{a<\frac{S_n-np}{\sqrt{npq}}\leqslant b\right\}=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx.$$

这也可表为: 当 $\Delta x>0$ 很小时, 有

$$P\left\{x\leqslant\frac{S_n-np}{\sqrt{npq}}\leqslant x+\Delta x\right\}=\frac{1}{\sqrt{2\pi}}\int_x^{x+\Delta x}e^{-\frac{u^2}{2}}du$$

$$\approx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Delta x.$$

当 $p=q=\frac{1}{2}$、$n=10000$ 时，上式化为

$$
\begin{aligned}
&P\{5000+50x\leqslant S_{10000}\leqslant 5000+50(x+\Delta x)\}\\
&\quad=F(5000+50(x+\Delta x)-F(5000+50x)\\
&\quad\approx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Delta x.
\end{aligned}
$$

这个近似，恰描绘了上面所说的直方图可用正态曲线逼近.

如果设想硬币出现正面的概率 $P(x_i=1)=p$ 随着试验的次数 n 变，这时把 p 记为 p_n 于是当 n 不同时，有

$$P\{S_n=k\}=C_n^k p_n^k(1-p_n)^{n-k},\quad k=0,\ 1,\ \cdots,\ n.$$

先考虑 $p_n=\frac{\alpha}{n}$ 的情形($\alpha>0$). 这时

$$P\{S_n=k\}=C_n^k\left(\frac{\alpha}{n}\right)^k\left(1-\frac{\alpha}{n}\right)^{n-k}.$$

由于 $\lim\limits_{n\to\infty}\left(1-\frac{\alpha}{n}\right)^n=e^{-\alpha}$，于是得

$$\lim_{n\to\infty}P\{S_n=0\}=\lim_{n\to\infty}\left(1-\frac{\alpha}{n}\right)^n=e^{-\alpha}.$$

但

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{P\{S_n=k+1\}}{P\{S_n=k\}}\\
&\quad=\lim_{n\to\infty}\frac{C_n^{k+1}\left(\frac{\alpha}{n}\right)^{k+1}\left(1-\frac{\alpha}{n}\right)^{n-k-1}}{C_n^k\left(\frac{\alpha}{n}\right)^k\left(1-\frac{\alpha}{n}\right)^{n-k}}=\frac{\alpha}{k+1},
\end{aligned}
$$

故由上式递推，得

$$
\begin{aligned}
&\lim_{n\to\infty}P\{S_n=1\}=\alpha\lim_{n\to\infty}P\{S_n=0\}=\alpha e^{-\alpha},\\
&\lim_{n\to\infty}P\{S_n=2\}=\frac{\alpha}{2}\lim_{n\to\infty}P\{S_n=1\}=\frac{\alpha^2}{2}e^{-\alpha},\\
&\vdots\\
&\lim_{n\to\infty}P\{S_n=k\}=\frac{\alpha}{k}\lim_{n\to\infty}P\{S_n=k-1\}=\frac{\alpha^k}{k!}e^{-\alpha},\\
&\vdots
\end{aligned}
$$

极限分布为

$$\frac{\alpha^k}{k!}e^{-\alpha}, \qquad k=0,\ 1,\ 2,\ \cdots\cdots.$$

这是以 α 为参数的普阿松分布，若随机变量 Y 的分布为：

$$P(Y=k)=e^{-\alpha}\frac{\alpha^k}{k!}, \qquad k=0,\ 1,\ 2\cdots\cdots,$$

容易算出 $EY=\alpha$, $\sigma^2(Y)=\alpha$, 可记为 $\pi(\alpha)$.

上面推导中假设了 $p_n=\frac{\alpha}{n}$, 事实上若把此假设放松为

$$\lim_{n\to\infty} np_n=\alpha,$$

极限分布仍是普阿松分布. 前面讲的"配对问题"情况正是 $p_n=\frac{1}{n}$.

二项分布、正态分布和阿松分布 是常用到的三个重要分布. 普阿松分布是许多实际随机现象的很适用的数学模型, 它在放射性分裂、电话接线、细胞中染色体的交换及细菌的分布等问题中都是很好的数学模型.

以上讨论了重复扔硬币的两种情形, 由于对 X_i 的分布的假设不同, S_n 的极限分布也不相同. 根据这两个特殊例子, 可提出一般的问题: 设 X_i (其中 $i=1,\ 2,\ \cdots\cdots$)为相互独立随机变量序列, S_n 为它们的部分和, 即 $S_n=\sum_{i=1}^{n}X_i$, 当 $n\to\infty$ 时 S_n 的分布将如何呢?这个问题称为"中心极限问题", 它是概率论中最重要的一类问题, 其中的定理有广泛的理论与实际意义.

把德莫阿-拉普拉斯定理里的条件 $p_n=p$ 取消, 且对每个 X_i 不必限制为相同的二值分布, 只假定 $EX_i=a_i$, $\sigma^2(X_i)=\sigma_i^2$, $i=1,\ 2,\ \cdots\cdots$. 这时如令 $s_n^2=\sum_{i=1}^{n}\sigma_i^2$, 考虑 S_n 的标准化

$$S_n^*=\frac{\sum_{i=1}^{n}(X_i-a_i)}{s_n}$$

$\left(当\ p_n=p\ 时,\ a_i=p,\ \sigma^2(X_i)=pq,\ s_n^2=npq,\ S_n^*=\frac{S_n-np}{\sqrt{npq}}\right)$.

李雅普诺夫(А. М. Ляпунов)改进了契比雪夫(П. Л. Чебышев)的矩法，用之给出了如下的定理：对一切 a、b，$-\infty<a<b<\infty$，若对某个正数 $\delta>0$，能成立 $\lim\limits_{n\to\infty}\dfrac{\sum\limits_{i=1}^{n}E|X_i-a_i|^{2+\delta}}{s_n^{2+\delta}}=0$，则有

$$\lim_{n\to\infty}P\{a\leqslant S_n^*\leqslant b\}=\int_a^b\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du.$$

此性质简称渐近正态性.

随着特征函数方法的引入，中心极限定理的研究得到了很快的发展．本世纪20年代，林德伯和莱维解决了诸随机变量 X_i(其中 $i=1, 2, \cdots\cdots$)为相同分布时的中心极限定理．设 $EX_i=a$，$\sigma^2(X_i)=\sigma^2<\infty$，则部分和的标准化 $S_n^*=\dfrac{S_n-na}{\sigma\sqrt{n}}$ 有渐近正态性.

1935年，林德伯和费勒又进一步对非相同分布的独立随机变量序列进行研究，他们得到了最一般的中心极限定理，使前人所得的结果都是它的推论．此结果叙述如下：设 $EX_i=a_i$，$\sigma^2(X_i)=\sigma_i^2$，则标准化部分和 S_n^* 为渐近正态且费勒条件

$$\lim_{n\to\infty}\frac{\max\limits_{1\leqslant i\leqslant n}\sigma_i^2}{s_n^2}=0$$

成立的充分必要条件是如下的林德伯条件成立：对任意 $\tau>0$，

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{i=1}^{n}\int_{|x-a_i|\geqslant\tau s_n}(x-a_i)^2dF_i(x)=0,$$

其中 $F_i(x)=P\{X_i\leqslant x\}$.

2. 中心极限定理的发展

此后中心极限定理的研究，是围绕以下几个方面进行的．减弱对随机变量序列的独立性的要求，而考虑具有某种相依性的随机变量序列；讨论向标准正态密度函数收敛的问题；及向正态分布收敛的速度及有关问题；以及大偏差理论等问题.

(i) **相依随机变量的中心极限定理** 这一问题仍是很活跃的

研究方向，其中讨论较多且获得实际应用的有 m **相依**随机变量序列，强平稳随机变量序列，鞅序列，**马尔可夫过程**及其泛函，以及各种类型的**统计量**序列．在相应的充分条件下，可以证明中心极限定理成立．

(ii) *局部极限定理* 向正态密度函数收敛的问题很早以前就有人讨论，但直至本世纪中期才得到一般性的结果．在棣莫弗-拉普拉斯定理的形成过程中，首先解决的是重复 n 次扔硬币时，正面出现次数 $S_n=k$ 的概率渐近于正态密度的问题，即在任意给定的有穷区间 (a, b) 中，对于满足 $a\leqslant\frac{S_n-np}{\sqrt{npq}}\leqslant b$ 的 S_n，一致地有

$$\lim_{n\to\infty}\frac{P(S_n=k)}{(npq)^{-\frac{1}{2}}\varphi(x_k)}=1.$$

其中 $x_k=\frac{k-np}{\sqrt{npq}}$，$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$．这个结果称为**棣莫弗-拉普拉斯局部极限定理**．此问题的推广是讨论取值为 $b+Nk$（其中 $N=0, \pm1, \pm2, \cdots$）的独立随机变量序列 (X_i) 相应的问题，即格点极限定理．对独立同分布情形，1948 格涅坚科给出了充分必要条件；对独立非同分布情形，于 50 年代也给出了充分条件，当独立随机变量序列 $\{X_n\}$ 的标准化和 S_n^* 有密度函数 $p_n(x)$ 时，讨论 $p_n(x)$ 向标准正态密度函数收敛的问题称为**局部极限定理**．格涅坚科于 1953 年对独立同分布的情形，给出了十分简洁的充分必要条件，即：当且仅当存在某个正整数 N，使 $p_N(x)$ 有界时，局部极限定理成立，即 $\lim\limits_{n\to\infty}\sup\limits_{x}\left|p_n(x)-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right|=0$．对非同分布的情形，彼得洛夫给出了局部极限定理成立的充分必要条件．

(iii) *收敛速度问题* 本世纪 40 年代，贝瑞和爱森先后研究了独立随机变量序列 (X_n) 的标准化部分和 S_n^* 的分布函数 $F_n(x)$ 向 $\Phi(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du$ 趋近的速度问题，他们的结果是：当 $EX_n=0$（其中 $n=1, 2, \cdots\cdots$），$EX_n^2<\infty$，$E|X_n|^3<\infty$ 时，令

$s_n^2=\sum_{i=1}^{\infty}EX_i^2$，则有

$$\sup_x|F_n(x)-\Phi(x)|\leqslant AL_n.$$

其中 $L_n=s_n^{-\frac{3}{2}}\sum_{i=1}^{n}E|X_i|^3$，$A$ 为常数．这一不等式给出了 $F_n(x)$ 向 $\Phi(x)$ 趋近的速度．对于这方面的研究，已进行得相当深入．

(iv) 大偏差定理　对独立同分布随机变量序列 (X_n)，若 $EX_i=0$，$EX_i^2=\sigma^2<\infty$，则对常数 $M>0$，有

$$\lim_{n\to\infty}\sup_{0<X<M}\frac{P(S_n^*>x)}{1-\Phi(x)}=1,\ \lim_{n\to\infty}\sup_{0<X<M}\frac{P(S_n^*\leqslant -x)}{\Phi(-x)}=1.$$

如果上述 M 不是常数，而是随着 n 而增大的 M_n，且 $M_n\nearrow\infty$（当 $n\to\infty$ 时），则与上述相类似的结果称为**大偏差定理**．确切地说，如果

$$\lim_{n\to\infty}\sup_{0<X<M_n}\frac{P(S_n^*>x)}{1-\Phi(x)}=1,\ \lim_{n\to\infty}\sup_{0<X<M_n}\frac{P(S_n^*<-x)}{\Phi(-x)}=1,$$

则称对 M_n 的大偏差定理成立．1938 年，克拉美在渐近展开的基础上证明：若存在正常数 H，使得当 $|t|<H$ 时，$Ee^{tX_1}<\infty$，则对 $M_n=0(\sqrt{n})\left(\text{即}\lim\limits_{n\to\infty}\frac{M_n}{\sqrt{n}}=0\right)$的大偏差定理成立．其后，林尼克给出了对 $M_n=bn^{\alpha-\frac{1}{2}}$ 的大偏差定理成立的充分必要条件，这里 b 为正常数，$\frac{1}{2}<\alpha<1$．大偏差定理在理论上很重要，而且，直到现在还被理论工作所重视，尚在不断地向深度发展．

3. 普遍极限定理

当我们仅考虑相同分布独立随机变量序列 (X_n) 时，当 $n\to\infty$，标准化部分和 S_n^* 的分布的极限分布是很小的一类分布．我们看到，如果 X_i 具有 $\delta(\geqslant 2)$ 阶矩：$E|X_i|^\delta<\infty$，那么 S_n^* 的极限分布就是正态分布．如果把条件 $E|X_i|^\delta<\infty$ 减弱为只对 $0<\delta\leqslant 2$ 时

成立，则 S_n^* 的极限分布包含很多分布，它们构成一个分布族，叫作稳定分布族，记为 $\mathscr{S}$ 族，正态分布是其中的一个成员．如果考虑不是相同分布的情形，而且假设 $\lim\limits_{n\to\infty}\max\limits_{1<i<n}P\{|X_i|>\varepsilon\}=0$ 对任意 $\varepsilon>0$ 成立（此要求是说在 S_n 中每个 X_i 取 $(-\varepsilon,\ \varepsilon)$ 之间值的概率接近于 1，这意味着在 S_n 中每个加项 X_i 所起的作用都很微小，此条件称为**无穷小条件**），那么 S_n^* 的极限分布族又可以扩大为 $\mathscr{L}$ 族．但 $\mathscr{L}$ 族还不包含很重要的普阿松分布．仔细考察从二项分布导出极限分布为普阿松分布的过程，可以看出作第 n 次伯努利试验时所考虑的 X_i（其中 $i=1,\ \cdots,\ n$）具有性质 $P(X_i=1)=p_n$，此处 p_n 依赖 n，于是作第 $n+1$ 次试验时，实际上我们所考虑的 $n+1$ 个随机变量的前 n 个不是 n 次时的随机变量，因为这时 $P(X_i=1)=p_{n+1}$，所以对应于第 n 次伯努利试验的随机变量，应如下编号才能有所区别：$X_{n,1},\ X_{n,2},\ \cdots,\ X_{n,n}$．换句话说，要想得到"部分和"的分布的极限为普阿松分布，必须考察随机变量阵列：$\{X_{n1},\ X_{n2},\cdots,X_{nk_n}\}$，$n=1,\ 2,\ \cdots$，其中当 $n\to\infty$ 时，$k_n\to\infty$ 而且每行中的随机变量相互独立，令 $S_n=\sum\limits_{k=1}^{k_n}X_{nk}$．普遍极限问题的提法是：对于适当选取的常数 A_n，随机变量 S_n-A_n 的分布的极限有哪些分布？即分布族是什么族？向分布族中的分布收敛的充分必要条件是什么？这些问题在本世纪 40 年代中期已获圆满解决．辛钦于 1937 年证明对独立随机变量阵列 $\{X_{n1},\ X_{n2},\ \cdots,\ X_{nk_n}\}$，$n=1,\ 2,\ \cdots$，$k_n\to\infty(n\to\infty)$，若每行的随机变量满足无穷小条件；对任意 $\varepsilon>0$，$\lim\limits_{n\to\infty}\max\limits_{1<k<k_n}P(|X_{nk}|>\varepsilon)=0$，则对适当的常数列 A_n，S_n-A_n 的可能的极限分布构成一个分布族，此族叫做**无穷可分分布族**（由于需要特征函数及其它高深数学知识才能讲清无穷可分分布族，这里就省略不讲了）．随后，1944 年格涅坚科给出了 S_n 的分布 $F_n(x)$ 收敛于此分布族中一个具体的无穷可分分布 $F(x)$ 的充分必要条件．我国数学家许宝騄也找到了充分必要条件．

由普遍极限定理可导出向正态分布、普阿松分布及退化于零的分布(即随机变量等于 0 的概率为 1 的分布)的最一般的条件. 例如满足无穷小条件的独立随机变量阵列的行的和向正态分布 $N(m, \sigma^2)$ 收敛的充分必要条件是:

(i) 对任意 $\varepsilon>0$,

$$\lim_{n\to\infty}\sum_{k=1}^{k_n} P(|X_{nk}|>\varepsilon)=0;$$

(ii) 存在某个 $\tau>0$, 使

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\left[\int_{|X|<\tau} x^2 dF_{nk}(x)-\left(\int_{|X|<\tau} x dF_{nk}(x)\right)^2\right]=\sigma^2;$$

(iii) 存在某个 $\tau>0$, 使

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}\int_{|X|<\tau} x dF_{nk}(x)=m.$$

此处 $F_{nk}(x)=P(X_{nk}\leqslant x)$. 这是最一般的中心极限定理, 林德伯-费勒定理可以由它推出.

应用中心极限定理, 可以得到**大数定理**(即极限分布为在 0 退化的分布). 大数定理验证了以频率解释概率的直观意义. 设事件 E 的概率为 p, 即 $P(E)=p$, 我们重复作实验, 定义随机变量序列 $\{X_n, n\geqslant 1\}$ 如下: $X_n=1$ 表示在第 n 次实验, 事件 E 发生, $X_n=0$ 表示 E 不发生, 于是 $P(X_i=1)=p$, $P(X_i=0)=1-p=q$, 这正是伯努利模型. 由中心极限定理知, 在 n 次试验中, E 发生的频数 $S_n=\sum_{i=1}^{n} X_i$ 的标准化随机变量的极限分布为正态, 即对任意 $l>0$, 有

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n-np}{\sqrt{npq}}\right|<l\right\}=\int_{-l}^{l}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du,$$

所以, 对任意给定的 $\varepsilon>0$, 只要把 n 取得足够大, 使得 $l<\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}$, 就有

$$\left\{\left|\frac{S_n-np}{\sqrt{npq}}\right|<l\right\}=\left\{\left|\frac{S_n-np}{n}\right|<l\frac{\sqrt{pq}}{\sqrt{n}}\right\}$$

$$\subset\left\{\left|\frac{S_n}{n}-p\right|<\varepsilon\right\},$$

于是
$$P\left\{\left|\frac{S_n-np}{\sqrt{npq}}\right|<l\right\}\leqslant P\left\{\left|\frac{S_n}{n}-p\right|<\varepsilon\right\}.$$

注意对任意小的 $\delta>0$, 可先选 l, 使得 $\int_{-l}^{l}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du>1-\delta$, 由此, 对任意给定的小 $\varepsilon>0$, 取 n 很大, 可使

$$P\left\{\left|\frac{S_n}{n}-p\right|<\varepsilon\right\}\geqslant P\left\{\left|\frac{S_n-np}{\sqrt{npq}}\right|<l\right\}$$

$$\doteq\int_{-l}^{l}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du>1-\delta.$$

这就是对任意给定的小 $\varepsilon>0$,

$$\lim_{n\to\infty}P\left\{\left|\frac{S_n}{n}-p\right|<\varepsilon\right\}=1.$$

此式说明频率 $\frac{S_n}{n}$ 与 p 之差的极限分布为在 0 退化的极限分布(即 $(-\varepsilon,\varepsilon)$ 的概率为 1, 无论 ε 如何小; 也即取 0 的概率为 1). 此外, 此式还说明以概率接近于 1 的程度(很可信的), 即如果 n 很大的话, 频率 $\frac{S_n}{n}$ 与真正概率 p 之差很小.

在讨论普遍极限定理的同时, 辛钦于 1936 年还考虑了独立随机变量序列 $\{X_n\}$ 的普遍极限问题, 这就是讨论对适当选取的 $B_n>0$ 与 A_n, 一般标准化部分和 $S_n^*\triangleq B_n^{-1}\sum_{i=1}^{n}X_i-A_n$ 的分布的极限分布及向极限分布的收敛条件, 在无穷小条件下, S_n^* 的极限分布构成一个分布族, 称为 $\mathscr{L}$ 族, 当然它是无穷可分分布族的子族. 1946 年, 莱维给出一个分布 $F(x)$ 是 $\mathscr{L}$ 族中的一个分布的充分必要条件. 随后, 格涅坚科又给出 S_n^* 的分布向 $\mathscr{L}$ 族中某个分布收敛的充分必要条件. 当随机变量序列 (X_n) 限定为独立同分布时,

S_n^* 的极限分布族更小一点，此族即本节开头所提的稳定分布族 $\mathscr{S}$. 1936 年，莱维与辛钦给出分布 $F(x)$为稳定分布的充分必要条件. 莱维、辛钦与费勒又分别给出了独立同分布为 $F_0(x)$ 的随机变量序列(X_n)服从中心极限定理的充分必要条件是

$$\lim_{M\to\infty}\frac{M^2\int_{|x|>M}dF_0(x)}{\int_{|X|<M}x^2dF_0(x)}=0.$$

此定理的推广，即向某个稳定分布(不限正态分布)收敛的充分必要条件问题，是由格涅坚科与杜叶勃林分别给出的.

极限定理是概率论的重要内容，也是数理统计的基础之一，其理论成果也比较完美. 长期以来，对于极限定理的研究所形成的概率论分析方法，至今仍影响着概率论的发展. 同时，由于讨论独立同分布(X_i)的部分和 $S_n=\sum\limits_{i=1}^{n}X_i$ 的某些连续泛函，如 $\max\limits_{1<k<n}S_k$ 的极限分布问题时，引起了随机过程的极限理论的研究，从而形成了不变原理，这里就不详细讨论了.

四、随机过程

1. 随机过程

随机过程就是一族随机变量 $(X_t, t\in T)$, T 叫做指标集，它可以是离散集$(t_i, i=1, 2, \cdots)$或$(0, 1, 2, \cdots)$，也可以是实轴$(-\infty, \infty)$上的任意区间或全实轴. 每个 X_t 的取值范围都是一样的，这个取值范围叫做状态空间，记为 S. S 也可以是离散集，也可以是连续集.

[例 1] 还是以重复扔硬币为例. 假定是无穷无尽地扔，于

是，所得结果可表为一随机变量无穷序列 $X_1, X_2, \cdots$，其中 X_i 表示第 i 次扔硬币的结果，$X_i=1$ 表示出现正面，$X_i=-1$ 表示出现反面．出现的概率各为 $\frac{1}{2}$．这种理想实验(当然是理想的，因为一个人连续扔，即使到死也未必能作完一次试验)的每一结果，是一串由 ± 1 组成的无穷序列，每一个这样的序列叫做这个随机过程的**轨道**(或**实现**)．过程$(X_n, n=1, 2, \cdots)$的状态空间 $S=(1, -1)$，参数集 $T=(1, 2, \cdots)$．

[例 2]　如上，我们考虑一串相互独立的取正数值的随机变量$(T_i, i=1, 2, \cdots)$，而且，每个 T_i 是相同分布的，其公共分布为以 $\alpha>0$ 为参数的指数分布(参看本书第 24 页中的例 3)：

$$P(T_i\leqslant t)=1-e^{-\alpha t}=\int_0^t \alpha e^{-\alpha t}dt, \quad t\geqslant 0$$

$$P(T_i>t)=e^{-\alpha t}, \quad t\geqslant 0$$

因为诸 T_i 相互独立，故对任意正整数 n 及非负实数 $t_1, t_2, \cdots, t_n$，有

$$P(T_i>t_i, i=1, \cdots, n)=P(T_1>t_1, T_2>t_2, \cdots, T_n>t_n)$$

$$=\prod_{i=1}^{n} P(T_i>t_i)=e^{-\alpha(t_1+\cdots+t_n)}.$$

这个模型可以作为电话台呼唤流或公路上汽车流的模型．令 $S_n=T_1+T_2+\cdots+T_n$，它表示第 n 次电话呼叫到达的时间(或第 n 辆车经过观察站的等待时间)．事件 $\{S_n\leqslant t\}$ 表示直到时刻 t 以前已有 n 次电话呼叫，另一种说法是：在时间区间$[0, t]$内电话呼叫总数至少为 n 次(后一种说法很有用)．如果令 $N_t=N(t)$ 表示在时间区间$[0, t]$内电话呼叫的次数，于是

$$\{S_n\leqslant t\}=\{N_t\geqslant n\}.$$

当然对每个 $t\geqslant 0$，N_t 也是随机变量．由此可见，从$(T_i, i=1,2,\cdots)$出发，我们得到随机过程$(N_t, t\geqslant 0)$，这里约定 $N_0=0$$(N_t, t\geqslant 0)$的状态空间 S 为$(0, 1, 2, \cdots)$，参数集 $T=[0, \infty)$．易知

$$(N_t=n)=(N_t\geqslant n)\setminus(N_t\geqslant n+1)$$
$$=(S_n\leqslant t)\setminus(S_{n+1}\leqslant t)=\{S_n\leqslant t<S_{n+1}\},$$

由此得知

$$P\{N_t=n\}=P(S_n\leqslant t)-P(S_{n+1}\leqslant t),\quad n=0,\ 1,\ 2,\ \cdots.$$

由计算,知

$$P(N_t=n)=\frac{(\alpha t)^n}{n!}e^{-\alpha t},\quad n=0,\ 1,\ 2,\ \cdots.$$

这个分布正是以 αt 为参数的普阿松分布 $\pi(\alpha t)$(参看本书第 31 页中心极限定理).

设随机变量 T 的分布为以 α 为参数的指数分布,于是对 s、$t>0$,有:

$$P(T>t+s|T>s)=\frac{P(T>t+s,\ T>s)}{P(T>s)}$$
$$=\frac{P(T>t+s)}{P(T>s)}=e^{-\alpha t}=P(T>t),$$

这种性质称为**无记忆性**,即 T 无论从那个时刻算起,它往后的条件分布,只和起始时刻到所要考察时刻之长有关,而与起始时刻无关.令 $N(s,\ s+t)=N_{s+t}-N_t$,$N(0,\ t)=N_t$,由上性质可以推出 $N(s,\ s+t)$ 的分布与 N_t 的分布相同.这种性质对于随机过程 N_t 来说,叫做**齐次性**.还可以证明,在不相交的时间区间 (s_1, s_1+t_1),$(s_2,\ s_2+t_2)$,$\cdots$ 上,随机变量 $N(s_1,\ s_1+t_1)$,$N(s_2,\ s_2+t_2)$,$\cdots$ 为相互独立的,而它们的分布分别为 $\pi(\alpha t_1)$,$\pi(\alpha t_2)$,$\cdots$.总之,随机过程 $(N_t,\ t\geqslant 0)$ 具有下列性质:

(i) $N_0=0$;

(ii) 对任意正整数 k 及不相交时间区间 $(s_i,\ s_i+t_i)$,增量 $N(t_i+s_i)-N(s_i)$ 为独立随机变量($i=1,\ \cdots,\ k$).此性质称为**独立增量性**;

(iii) 对每个 $s\geqslant 0$、$t\geqslant 0$,增量 $N(t+s)-N(s)$ 的分布为普阿松分布 $\pi(\alpha t)$,此性质称为**齐次性**.

具有上述性质的过程,称为**普阿松过程**.它的每个轨道为跳

步等于 1 的上升阶梯函数，如图 11，是一个轨道的图形：

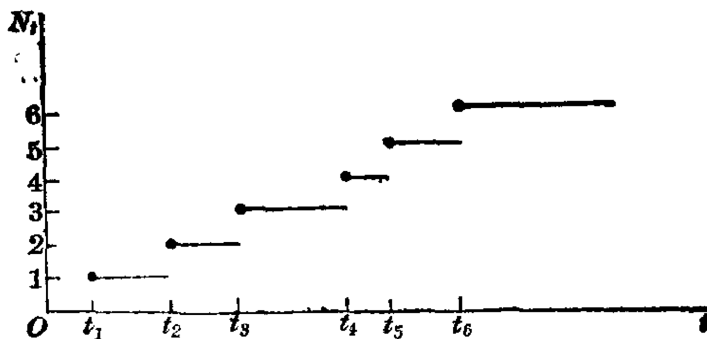


图 11

[例 3]（布朗运动） 1828 年英国生物学家布朗在显微镜下观察到小质点在液体中悬浮时，表现出一种奇怪而飘忽不定的游动，这种游动后来被称为"布朗运动"，它是由于周围液体分子的连续碰撞而引起的. 爱因斯坦和斯莫路稠夫斯基对此现象在理论上用概率论进行分析. 当然，布朗质点是在三维空间中游动，但我们可以研究它在坐标轴上的投影的游动. 因为质点每秒受到非常多的碰撞，我们要缩小时间单位，同时也缩小长度单位，才能方便准确地描述质点运动. 设新的时间单位很短很短，暂记为 δ，它表示某个质点的两次相继碰撞的间隔时间，于是质点在 $t=0$ 时，处在原点，在 t 时间内，共遇到 $\frac{t}{\delta}$ 次碰撞. 每次碰撞使质点向左或向右位移，位移之长度等于 $\sqrt{D\delta}$，其中 D 叫作扩散系数，于是第 k 次位移 ξ_k 是随机变量，可用伯努利试验作其近似模型：

$$P(\xi_k=\sqrt{D\delta})=P(\xi_k=-\sqrt{D\delta})$$
$$=\frac{1}{2}. \qquad k=1,\ 2,\ \cdots.$$

易知 $E\xi_n=0$，$\sigma^2(\xi_k)=D\delta$. 令 $X_0\equiv 0$，$X_n=X_0+\xi_1+\cdots+\xi_n$，布朗质点在时刻 t 的位置约等于质点受 $\frac{t}{\delta}$ 次碰撞后的位置：

$$X_t=\sum_{k=1}^{t/\delta}\xi_k.$$

易知 $EX_t=0$, $\sigma^2(X_t)=\frac{t}{\delta}D\delta=Dt$. 当 t 固定，令 $\delta\to0$ 时，由棣莫弗-拉普拉斯中心极限定理知 X_t 的分布为 $N(0, Dt)$，其密度为 $\frac{1}{\sqrt{2\pi Dt}}e^{-\frac{X^2}{2Dt}}$.

扩散系数 $D=\frac{2RT}{Nf}$，其中 R 为理想气体常数，T 为绝对温度，f 为摩擦系数，与液体粘度及质点性质有关，而 N 为物质一克分子所含分子的个数，即阿伏伽德罗数，总之，布朗运动过程 X_t 具有下述性质：

(i) $X_0=0$;

(ii) 在有穷个不相交的时间区间 (s_i, s_i+t_i) 上过程增量 $X_{s_i+t_i}-X_{s_i}$ 为相互独立的随机变量($i=1, 2, \cdots, k$);

(iii) 对每个 $s\geqslant0$ 及 $t\geqslant0$, $X_{s+t}-X_s$ 的分布密度为正态分布，其密度不依赖于 s，等于 $\frac{1}{\sqrt{2\pi Dt}}e^{-\frac{x^2}{2Dt}}$.

此外，几乎所有的轨道都是 t 的连续函数，但是处处不可微，即这些连续轨道在每个点上没有切线．我们可以对布朗运动每隔一定的短时间所观察到的质点投影到一张平面上，画出一个轨道的投影图形．可以设想，在$[0, t]$内大量观察布朗质点，把水平位移 X_t 记录下来，把它们画成直方图，此图与正态曲线近似．利用这个办法可以估计 D，从而可以估计阿伏伽德罗数 N．本世纪初，伯林就用这种思想作了大量试验，求得了 N，从而获得诺贝尔物理学奖．从布朗运动试验而成功地定出阿伏伽德罗数，是科学史上一个划时代的事例，因为有了它，才使姗姗来迟的原子理论得到承认，从而推动了物理学的大大发展．这正符合拉普拉斯的预言："值得注意的是，从考虑赌博问题而源起的一门科学，将必成为人类知识宝库里最重要的课题"．由此例，还知随机过程所取的值可以是三维的．多维的布朗运动有一个依赖于维数的有趣的性质，就是常返性．当维数为 1 与 2 时，布朗运动是常返的，即它从任一

点出发，几乎所有轨道或迟或早地返回出发点，但当维数高于2时，此性质不再保留，即几乎所有轨道都远离出发点.

2. 随机过程的分类

根据随机过程$(X_t, t\in T)$中诸随机变量间的不同的相依关系，可以把随机过程进行分类：

(1) **独立增量过程** 在任何一组不相交区间上，过程的增量都是相互独立的过程，叫做**独立增量过程**. 即对任意正整数n及任意$t_0<t_1<t_2<\cdots$, $t_i\in T$, $i=0, 1, 2, \cdots n$. 增量$X_{t_i}-X_{t_{i-1}}$ ($i=1, 2, \cdots, n$)及X_{t_0}是相互独立的. 如果增量$X_{t+h}-X_h$的分布只与t有关，而与h无关，则过程为齐次增量的过程. 前面所提的普阿松过程及布朗运动都是齐次的独立增量过程.

(2) **平稳过程** 过程的概率特性不随时间推移而变化的随机过程叫做**平稳过程**. 例如一台稳定工作的纺纱机在时刻t纺出纱的直径X_t是随机变量，但在时刻$t_1, t_2, \cdots, t_n$时纱的直径$(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$与在时刻$t_1+\tau, t_2+\tau, \cdots, t_n+\tau$时$(\tau>0)$纱的直径$(X_{t_1+\tau}, X_{t_2+\tau}, \cdots, X_{t_n+\tau})$是无统计差别的，即这两组随机变量有相同的$n$维分布. 这时，随机过程$X_t$叫做**严平稳**. 如果二阶绝对矩$EX_t^2<\infty$，而且对任意$t$、$\tau$，均值$EX_t=m$(常数)、协方差$E[(X_{t+\tau}-m)(X_t-m)]=\Gamma_\tau$与$t$无关，则称$X_t$为**宽平稳的**. Γ_t称为过程的协方差函数. 一个严平稳过程，如果它的二阶矩有穷，则它一定也是宽平稳的.

平稳过程的基本理论是本世纪三十至四十年代建立和发展起来的，但直到现在，关于平稳过程中离散参数的过程"**时间序列**"的研究还是很活跃的. 平稳过程理论在无线电技术和自动控制等领域有着广泛的应用，是信号分析、滤波、预测理论及控制论等应用学科的重要工具.

(3) **马尔可夫过程** 人们在实际中常遇到具有下述特性的随

机过程：在已知它目前的状态（"现在"）的条件下，它未来的演变（"将来"）不依赖以往的演变（"过去"）．这种在已知"现在"的条件下，"将来"与"过去"独立的特性称为**马尔可夫性**，这概念是俄国数学家马尔可夫（A. A. Марков）于1907年提出来的．在一般情形下，已知"现在"，对"将来"演变的可能随着对"过去"的了解愈多愈能有所补益的，但马尔可夫性则无需"过去"知识的增多．用数学语言来表达，对于随机过程 X_t，如果对任意 $t_1<t_2<\cdots<t_n<t$，有

$$P\{a<X_t\leqslant b \mid X_{t_1}=x_1,\ X_{t_2}=x_2,\ \cdots,\ X_{t_n}=x_n\}$$
$$=P\{a<X_t\leqslant b \mid X_{t_n}=x_n\},$$

这个随机过程 X_t 便叫做**马尔可夫过程**．如果取状态空间 $S=(0,\ 1,\ 2,\ \cdots)$，参数空间 $T=(0,\ 1,\ 2,\ \cdots)$，则马尔可夫过程记为 (X_n)，称之为**马尔可夫链**，即对任意 $0\leqslant n_1<n_2<\cdots<n_l<m$，$n>0$ 及非负整数 $i_1,\ i_2,\ \cdots,\ i_l,\ i,\ j$，有

$$P\{X_{n+m}=j \mid X_{n_1}=i_1,\ \cdots,\ X_{n_l}=i_l,\ X_m=i\}$$
$$=P\{X_{n+m}=j \mid X_m=i\}.$$

如果上式右方与 m 无关，则称 (X_n) 为**齐次马尔可夫链**．此时，上式右方是马尔可夫链从 i 出发经 n 步转移到 j 的条件概率，称之为 n 步转移概率，记为 $p_{ij}^{(n)}$．对马尔可夫链，人们只要知道 $(p_{ij}^{(n)})$，i、$j\in S$，就可知链的 n 步演变的规律．由于从 i 出发经 $n+m$ 步转移到 j 必是从 i 出发经 n 步转移到 S 中的任意状态 k，然后再从 k 经 m 步转移到 j（这时已知状态 k，故经 m 步时，其条件概率与过去无关），因此有

$$p_{ij}^{(n+m)}=\sum_{k\in S} p_{ik}^{(n)} p_{kj}^{(m)},$$

这就是切普曼-柯尔莫哥洛夫方程．根据此方程，任意步转移概率可通过一步转移概率计算出来，因此，只需知道一步转移概率 p_{ij}，就掌握了这个马尔可夫链的转移规律了．

［例］ 设 $(\xi_n,\ n\geqslant 0)$ 为相互独立随机变量序列，$\xi_0\equiv 0$，$S=$

$(\cdots -2, -1, 0, 1, 2, \cdots)$. 定义随机过程 $X_n = \sum_{k=0}^{\infty} \xi_k, n \geqslant 0$. 因为 $X_{n+1} = X_n + \xi_{n+1}$, 而且 ξ_{n+1} 与 $X_0, X_1, \cdots, X_n$ 相互独立. 令 A 表示与 $X_0, X_1, \cdots, X_{n-1}$ 有关的任何事件, 则有

$$
\begin{aligned}
P(X_{n+1}=j | A;\ X_n = i) &= P(\xi_{n+1} = j - i | A;\ X_n = i) \\
&= P(\xi_{n+1} = j - i) \\
&= P(X_{n+1} = j | X_n = i) = p_{ij}.
\end{aligned}
$$

由此可知, 部分和 X_n 是马尔可夫链. 特别, 当 $P(\xi_i = 1) = p$、$P(\xi_i = -1) = 1 - p = q$ 时, X_n 的一步转移概率 p_{ij} 如下:

$$
p_{ij} = \begin{cases} p, & \text{若 } j = i+1; \\ q, & \text{若 } j = i-1; \\ 0, & \text{若 } j \neq i+1 \text{ 及 } i-1. \end{cases}
$$

这个模型叫做**自由随机游动**, 即以前讨论的伯努利模型.

关于马尔可夫过程的理论研究, 1931 年柯尔莫哥洛夫发表了《概率论中的解析方法》, 此文首先把**微分方程**等分析方法应用于此类过程, 奠定了它的理论基础. 其后, 又逐渐反过来以马尔可夫过程理论来研究微分方程的问题. 1951 年前后, 伊藤清在莱维及伯恩斯坦等人的工作基础上, 建立了**随机微分方程**的理论, 这为研究马尔可夫过程开辟了新的研究方法. 1954 年前后, 费勒将泛函分析中的**半群**方法引入马尔可夫过程的研究中, 邓肯等并赋予它以概率意义(如**特征算子**等). 本世纪五十年代初, 角谷静夫和杜勃等发现了布朗运动(也是特殊的马尔可夫过程)与偏微分方程中的狄利克雷问题的关系. 其后, 亨特研究了相当一般的马尔可夫过程(亨特过程)与**位势**间的关系. 目前, **流形上的马尔可夫过程**、马尔可夫场等, 都是正待深入研究的领域.

前面讲的普阿松过程也是马尔可夫过程, 它的轨道是取非负整数的增函数, 实际上它是连续时间的马尔可夫链. 具有连续轨道的马尔可夫过程叫做**扩散过程**, 布朗运动就是扩散过程的特例. 此外, 非常有用的**分支过程**及**生灭过程**等等, 都具有马尔可夫性.

(4) **鞅** 设 X_t 为实数值的随机过程，若对一切 $t\in T$，$E|X_t|<\infty$，且当 $t_1<t_2<\cdots<t_n<t$ 时，有

$$E(X_t|X_{t_1}=a,\ \cdots,\ X_{t_n}=a_n)=a_n,$$

此处 $a_1, a_2, \cdots, a_n$ 为相应随机变量的可能值，则称 X_t 为鞅. 鞅可以看作成公平赌博的适宜的模型. 把 X_t 看作一场赌博中某个局中人在时刻 t 的本钱，$X_{t_1}=a_1, \cdots, X_{t_n}=a_n$ 看作是此人到时刻 t_n 为止积累起来的经验，于是鞅的性质表明，不论他在时刻 t_n 以后的赌博中如何利用已得的经验"$X_{t_1}=a_1, \cdots, X_{t_n}=a_n$"，他所能期望于未来时刻 t 的本钱仍只能是 $X_{t_n}=a_n$，这是赌博公平性的体现. 均值为 0 的独立随机变量序列的部分和、布朗运动和均值为零的独立增量过程都是鞅的典型例子.

莱维等人早在 1935 年就发表了一些孕育着鞅论的工作. 1939 年，维勒首次采用"鞅"这个名称. 但对鞅进行系统研究并使之成为随机过程论的一个重要分支的，则归功于杜勃以及后来解决上鞅分解问题的数学家梅耶. 鞅论本身的研究固然是很重要的，但它作为工具更为重要，例如随机积分的研究就不能离开对鞅的随机积分. 此外，鞅论方法对解决分析(特别是调和分析)中的问题也是有力的工具.

(5) **点过程** 很多随机现象，它们发生的时刻或地点可以用某一空间上的点来表示，例如服务台前顾客的到来时刻，真空管阴极电子的放射时刻可表为实轴上的随机点. 又如天空中某一区域星体的分布，核医疗中放射性示踪物质在人体各器官的出现部位，都可用三维空间的点表示. 考虑路灯的替换更新问题，灯泡的寿命是一个随机变量 T，设在 0 时装上灯泡，在 T_1 时烧坏，当即换上一个，这个灯泡在 T_1+T_2 时又坏了，再换上一个，依此下去，……. 令 $S_n=T_1+T_2+\cdots+T_n$，其中每个 T_i 的分布都与 T 的分布一样，当然 $T_1, T_2, \cdots, T_n$ 相互独立. 随机变量序列 $S_1, S_2, \cdots, S_n$ 表示灯泡坏了(随机事件)的时刻，所以 (S_n) 是直线上的一些随机点，故为点过程(一维点过程). 如本书第 39 页上的例

2, 可以用 N_t 表示[0, t]区间内灯泡换上的个数(即随机事件发生的次数), 于是有:

$$N_t=n,\quad 当\ S_n\leqslant t<S_{n+1},\ n=0,\ 1,\ 2,\ \cdots$$

过程 N_t 是计数过程的特例, 称为"更新过程". 更一般的点过程的定义如下: 设 n 维空间 $\mathbf{R}^n$ 的子集用 B、B_1、B_2、$\cdots$ 表示, 以 $\mathbf{R}^n$ 中子集为指标的随机变量族 $\{N(B),\ B$ 为 $\mathbf{R}^n$ 中的子集$\}$ 称为**点过程**. 如果 $N(\cdot)$ 的状态集(即可能值)为 $\{0,\ 1,\ 2,\ \cdots,\ \infty\}$, $N(B)$ 就表示某随机现象在子集 B 内发生的次数. 因为 $N(A)$ 是起计数作用的, 故对不相交子集 B_1、B_2, 应有:

$$N(B_1+B_2)=N(B_1)+N(B_2).$$

此外, 当然还需假设 $N(\varnothing)=0$, 这时称为 n **维点过程**.

本世纪 60 年代以前, 点过程的研究着重于一维情形, 内容多为考虑普阿松过程的种种推广. 以后逐渐扩充到多维及更一般的空间的随机点分布问题, 并与迅速发展的随机测度及鞅论相结合, 成为当前活跃的一个分支. 点过程在随机服务系统、交通运输、物理学、生态学、神经生理学、传染病学等等领域有广泛的应用.

3. 随机游动与调和函数

作为介绍概率论这一数学分支的概貌的结束, 我们用随机游动与电网络之间的关系, 来说明随机问题与非随机问题之间在方法上的互通性.

先考虑在直线上的随机游动: 一个质点在直线上一步一步的走动, 每一步可以是向左或向右走一步, 向左、向右走的概率各等于 $\frac{1}{2}$, 点游动到 0, 或 n 时, 不再游动. 如下图:



图 12

点从 $x(1\leqslant x\leqslant n-1)$ 出发，求此点游到 n 的概率 $p(x)$. 这个问题也可以用赌博语言叙述：甲乙两人以元为单位赌钱，两人总共有的本钱之和等于 n，每赌一盘甲赢乙1元，其概率为 $\frac{1}{2}$，甲输乙1元，其概率为 $\frac{1}{2}$；一盘一盘的赌下去，直至甲的本钱为0（赌光），或为 n（全赢）为止. 这时，$p(x)$ 即是甲原有本钱 x 而全赢的概率.

因为 X 可以是 $0, 1, \cdots, n$ 中的任一个数，故 $p(x)$ 有下列性质：

（i）$p(0)=0$;

（ii）$p(n)=1$;

（iii）$p(x)=\frac{1}{2}p(x-1)+\frac{1}{2}p(x+1)$，$1\leqslant x\leqslant n-1$.

因为当我们考虑质点在 x 时走一步后所发生的情况时，它以 $\frac{1}{2}$ 的概率向右或左走一步而到达 $X+1$ 或 $X-1$. 在向右的条件下，它到 n 的概率将是 $p(x+1)$，同样，在向左的条件下，它到 n 的概率是 $p(x-1)$. 于是总的概率 $p(x)$ 将等于(iii)的右端. (iii)是关于 x 的差分方程，(i)、(ii)是边界条件. 由(iii)知

$$p(x)-p(x-1)=p(x+1)-p(x),\quad 1\leqslant x\leqslant n-1,$$

但由(i)、(ii)知

$$1=p(n)-p(0)=\sum_{i=1}^{n}[p(i)-p(i-1)],$$

故

$$p(i)-p(i-1)=\frac{1}{n},\qquad i=1, 2, \cdots, n,$$

于是

$$p(i)=\frac{i}{n},\qquad i=1, 2, \cdots, n-1.$$

所以

$$p(x)=\frac{x}{n}.$$

现在让我们看另一个显然不同的问题，我们把0到 n 每二个邻点之间用相同的电阻 R 串联起来，而把两个端点0、n 用1伏电

压联接起来，0 点接地，如图 13. 令 $\nu(x)$ 记 $x=0, 1, \cdots, n$ 点上的电压，现在我们要求 $\nu(x)$. 这里当然有 $\nu(0)=0$ 与 $\nu(n)=1$，于是函数 $\nu(x)$ 满足上列条件(i)与(ii).

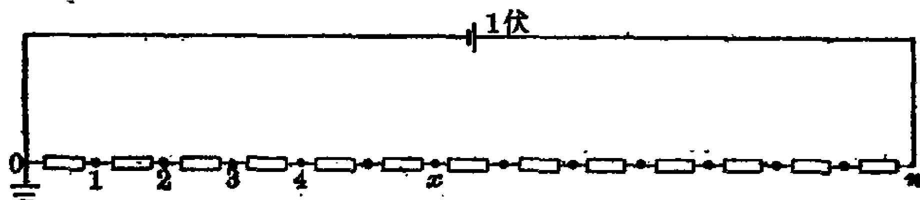


图 13

由基尔霍夫定律，流入 x 的电流应等于从它流出的电流. 又由欧姆定律，知两邻点 x、y 用电阻 R 相联时，则由 x 流入 y 的电流 $i_{xy}=\dfrac{\nu(x)-\nu(y)}{R}$. 于是，对 $x=1、2、\cdots、n-1$，有

$$\frac{\nu(x-1)-\nu(x)}{R}+\frac{\nu(x+1)-\nu(x)}{R}=0,$$

由此得

$$\nu(x)=\frac{1}{2}\nu(x+1)+\frac{1}{2}\nu(x-1), \quad x=1, 2, \cdots, n-1.$$

由此可见，$\nu(x)$ 也满足上述的差分方程(iii). 所以，$p(x)$ 与 $\nu(x)$ 都满足(i)、(ii)、(iii). 如果能证明满足(i)、(ii)、(iii)的解是唯一的，那么，就可以看出随机游动与电网络之间的互通性. 对此简单情形，我们很易证明. 在证明它以前先引进一些分析学的概念. 记 $S=\{0, 1, \cdots, n\}$，边界点 $\{0, n\}$ 记为 B，内点集 $\{1, 2, \cdots, n-1\}$ 记为 D. 定义于 S 上的函数 $f(x)$ 如果在 D 上每一点都有下列的平均性质：

$$f(x)=\frac{f(x-1)+f(x+1)}{2},$$

就称为**调和函数**. 上述的 $p(x)$、$\nu(x)$ 都是 S 上的调和函数，而且 $p(0)=\nu(0)=0$; $p(n)=\nu(n)=1$. 所以，问题化为求一个边界值已知的调和函数问题，这个问题称为**狄利克雷问题**. 它的唯一性原则说明，**不能有二个不同的调和函数具有相同的边界值**. 由此

可知 $p(x)=\nu(x)$. 下面我们证明唯一性原则，为此先证明调和函数的极大值原则，它与唯一性原则的关系正像微积分里洛尔定理与中值定理之间的关系一样.

极大值原则：定义于 S 上的调和函数 $f(x)$ 在 S 的边界 $B=\{0, n\}$ 上取其极大值 M 与极小值 m. 可用反证法. 设若不然，设 f 在 $x\in D$ 上取 $f(x)=M$，因为 $f(x)$ 是 $f(x-1)$ 与 $f(x+1)$ 的平均且 M 极大，故 $f(x-1)=f(x+1)=M$. 若 $x-1$ 仍在 D 内，同样可知 $f(x-2)=M$；依此类推，必得 $f(0)=M$. 对 m 可用同样的方法处理.

唯一性原则：若 $f(x)$ 与 $g(x)$ 为 S 上的调和函数，使得 $f(x)=g(x)$，$x\in B$，则 $f(x)=g(x)$ 对一切 x 成立. 令 $h(x)=f(x)-g(x)$. 若 x 为内点，则有

$$\begin{aligned}&\frac{h(x-1)+h(x+1)}{2}\\&=\frac{f(x-1)+f(x+1)}{2}-\frac{g(x-1)+g(x+1)}{2}\\&=f(x)-g(x)=h(x),\end{aligned}$$

故 $h(x)$ 为调和函数. 但当 $x\in B=\{0, n\}$ 时，$h(x)=0$，于是，由极大值原则知 h 的极大极小值都是 0. 故 $h(x)=0$ 对一切 x 成立. 即对一切 x，有 $f(x)=g(x)$.

最后让我们考察二维空间的随机游动. 图 14 描述这样的游动：黑点表示边界点，标有 E 的边界点表示"出口"，标有 F 的边界点表示陷井. 标有"×"的点为内点. 求游动者从内点 x 出发，在它未掉下陷井之前可以出去的概率 $p(x)$. 游动者从 $x=(a, b)$ 走一步时，可能到达上、下、左、右格子点的概率各为 $\frac{1}{4}$，上、下、左、右格子点即是点：$(a, b+1)$、$(a, b-1)$、$(a-1, b)$、$(a+1, b)$，若游动者游到边界点，就停止游动，此问题相应的电网络问题可表示为图 15. 其中所有边界点 F 相联接地，所有 E 点相连，最后的 E、F 用 1 伏电压联接. 问题是求在内点 x 处的电压 $\nu(x)$.

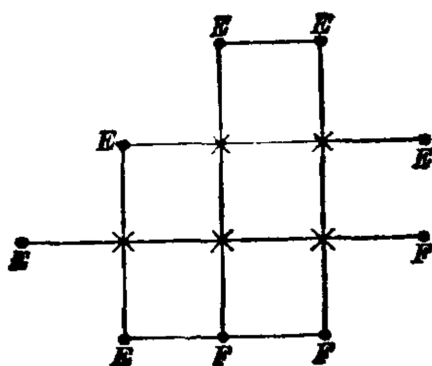


图 14

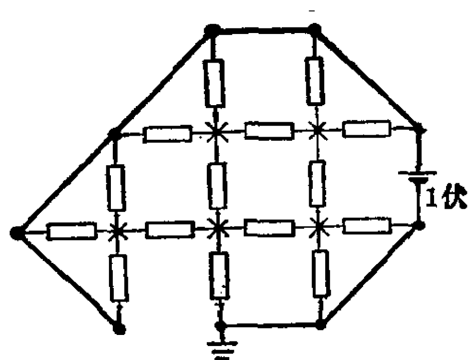


图 15

现在我们定义平面上格子点上的调和函数. 令 $S=D+B$, D、B 为不相交的格子点集合, 每个 D 上的格子点有上、下、左、右四个属于 S 的格点为其邻点, 而每个 B 的四个邻点中至少有一个点属于 D. 此外, 假定 S 中的点有连通性, 即对 S 中的任意两个点 P 及 Q, 必可找到一串属于 D 的点 $P_j(j=1, 2, \cdots, n)$ 使得 $P, P_1, P_2, \cdots, P_n, Q$ 为一条通路. 此时 D 为 S 的内点集, B 为 S 的边界点集. 定义于 S 上的函数 $f(x, y)$, 如果对每个内点 (a, b) 有下列平均性:

$$f(a, b)=\frac{1}{4}\{f(a+1, b)+f(a-1, b)+f(a, b+1)+f(a, b-1)\},$$

就称它为调和函数. 注意, 这时我们对 f 在边界点上之值未加限制. 欲证 $p(x)=v(x)$, 如同一维情形一样, 证 $p(x)$ 为调和函数应考虑一步之后的所有四个相邻点; 证 $v(x)$ 也为调和函数可用基尔霍夫定律, 流入点 $x=(a, b)$ 的电流为:

$$\frac{v(a+1, b)-v(a, b)}{R}+\frac{v(a-1, b)-v(a, b)}{R}+\frac{v(a, b+1)-v(a, b)}{R}+\frac{v(a, b-1)-v(a, b)}{R}=0,$$

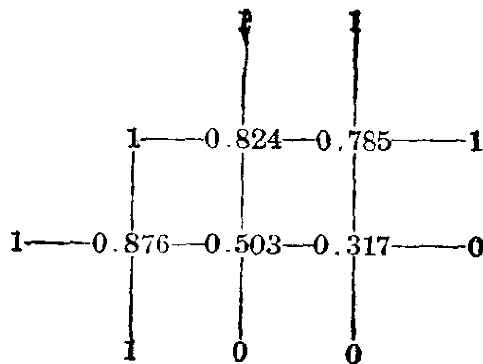
于是得

$$\nu(a, b)=\frac{1}{4}\{\nu(a+1, b)+\nu(a-1, b)+\nu(a, b+1)+\nu(a, b-1)\},$$

于是 $p(x)$、$\nu(x)$ 是具有相同边界值的调和函数，在每个 E 处等于 1，在每个 F 处为 0. 和一维情形一样，可证明极大值原则及唯一性原则也成立.

求解这个二维空间狄利克雷问题并不很简单，在给具体解法之前，让我们考虑两种求近似解的方法：

(i) 随机模拟法　用赌博语言来讲，狄利克雷问题的解，就是求下述赌博中局中人最终赢得的钱：局中人从 x 出发进行随机游动，直到他达到某一个边界点，他就赢得 $f(y)$ 个钱，如果 y 是他第一次所达到的边界点，这时 $f(y)=+1$ 或 0，当 y 为 E 类点或 F 类点. 于是，为求 $f(x)$，我们可以从 x 出发，重复很多次随机游动，因之得到很多个最终赢额(1 或 0)的平均值，由大数定律可知这个平均值可以作为 $f(x)$ 的相当精确的估计. 下面的图形就是作了 10000 次上述随机游动的计算结果：



(ii) 狄利克雷原问题·松弛法　上面所述的狄利克雷问题已并不是原来的问题，而是原来问题的离散说法. 原来的问题是考虑温度在连续介质中的分布问题. 典型的例子如下：设有一薄的方金属片，从当中挖去一个小方形. 在薄片(图 16)的内边上的温度为 0，而在外边上的温度为 1. 问题是求这金属薄片中一个点上的温度. 若以 $u(x, y)$ 表示在点 (x, y) 的温度，于是 $u(x, y)$ 满足拉普拉斯微分方程：

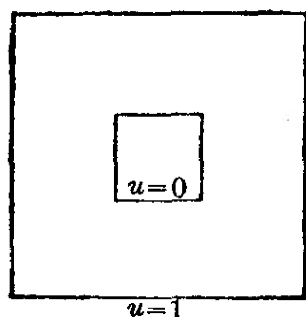


图 16

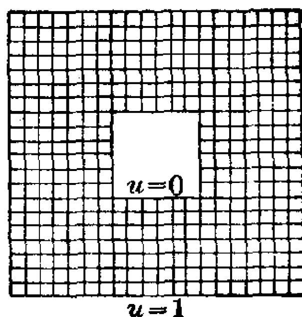
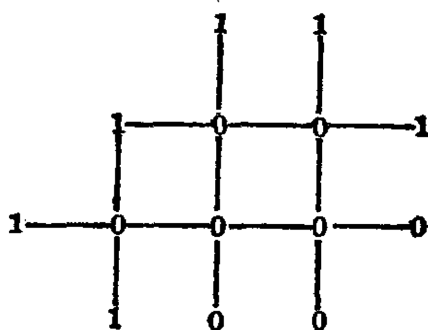


图 17

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0.$$

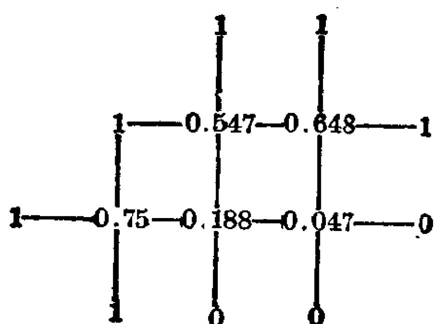
满足此方程的二元函数称为调和函数，它具如下的性质：在(x, y)处以它为圆心画任意在阴影中的圆，则$u(x, y)$等于u在圆周上所有诸值的平均．所谓狄利克雷问题，就是求调和函数$u(x, y)$，使它在内边界上等于0而在外边界上等于1．上述的离散型问题与此问题完全相应．用松弛法求原狄利克雷问题的近似解就是把原问题的连续介质代以格子点(如图17)，然后求一个调和函数(离散情形的)，而且在边界上取给定的值．做法是根据调和函数的平均性，即在一个内点函数之值等于四个邻点上之值的平均．首先把相应的边界值记下，然后对每个内点上给函数一个初值，例如在我们的离散问题中，对每个内点先给函数之值为0：



作一次相邻的平均得如下内点之值(四舍五入)：

$$\frac{1}{4}(1+1+1+0)=0.75,$$

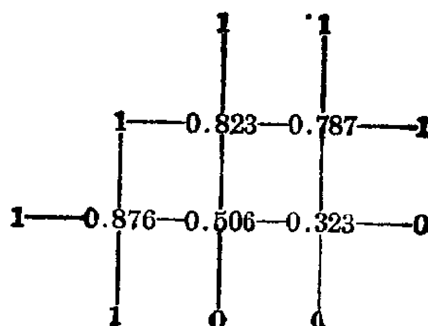
$$\frac{1}{4}(0.75+0+0+0)=0.1875,$$



$$\frac{1}{4}(0.1875+1+1+0)$$
$$=0.5469,$$
$$\frac{1}{4}(0.1875+0+0+0)$$
$$=0.0469,$$

$$\frac{1}{4}(0.0469+0.5469+1+1)=0.64844.$$

依此平均法再往下作，作到第 8 次得



下面我们给出确切解．内点用 a, b, c, d, e 表示，由平均性可知：

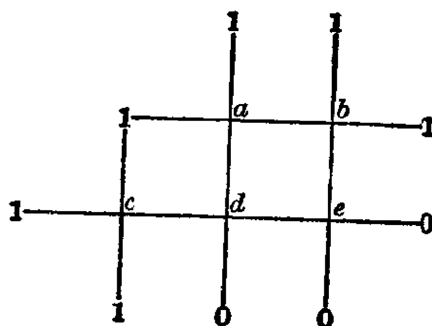
$$u_a=\frac{1}{4}(u_b+u_d+2),$$

$$u_b=\frac{1}{4}(u_a+u_e+2),$$

$$u_c=\frac{1}{4}(u_d+3),$$

$$u_d=\frac{1}{4}(u_a+u_c+u_e),$$

$$u_e=\frac{1}{4}(u_b+u_d).$$



解此联立方程组，得 $u_a=0.823$，$u_b=0.787$，$u_c=0.876$，$u_d=0.506$，$u_e=0.323$.

为了比较，列表如下：

	u_a	u_b	u_c	u_d	u_e
计　算:	0.823	0.787	0.876	0.506	0.323
随机模拟: (10000 次)	0.824	0.785	0.876	0.503	0.317
松 弛 法: (8 次)	0.823	0.787	0.876	0.506	0.323

在上面提到的连续情形中,当二元函数 $u(x,y)$ 在 $\mathbf{R}^2$ 中一个开集合 D 上满足拉普拉斯方程时, 它就叫做调和函数. 可以证明这个定义等价于如下的定义: $u(x,y)$ 在 D 上连续有界, 且有平均性, 则它是调和函数. 古典的狄利克雷问题如下: 设 D 为 $\mathbf{R}^2$ 中开集, 给一个在 D 的边界上的有界连续函数 f, 求一函数使它在 D 上为调和函数, 在 D+(D 的边界)上为连续, 在 D 的边界上等于 f. 可以想到用概率方法解此问题所要用到的手段是借助于布朗运动, 因为布朗运动可以用随机游动去逼近, 而随机游动在解决离散情形的狄利克雷问题时是概率方法所用到的手段. 不只如此, 还可以以布朗运动解决薛定谔方程的狄利克雷边界值问题, 这里就不深入展开了.